GENETIC AND PHENOTYPIC EVOLUTION OF BREAST CANCER THROUGH YEARS OF TREATMENT

by

Samuel Brady

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Biomedical Informatics

The University of Utah

August 2017

The University of Utah Graduate School

STATEMENT OF THESIS APPROVAL

The thesis of **Samuel Brady**

has been approved by the following supervisory committee members:

Andrea Bild	, Chair	**3/21/2017** Date Approved
Aaron Quinlan	, Member	**3/21/2017** Date Approved
Younghee Lee	, Member	**3/21/2017** Date Approved

and by **Wendy Chapman** , Chair/Dean of

the Department/College/School of **Biomedical Informatics**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Acquired drug resistance is a frequent challenge in breast cancer. A tumor may initially respond to chemotherapy, but later become resistant and relapse. This is largely due to intra-tumoral heterogeneity; tumor genetic subclones with higher fitness in response to chemotherapy survive, and the patient's cancer becomes drug-resistant. In additional, non-genetic phenotypic alterations in response to chemotherapy may promote drug resistance. How breast cancers evolve to become drug-resistant is unclear. A better understanding of how this occurs could lead to alternative therapeutic regimens to treat drug-resistant breast cancer or prevent its development. To address this problem, we performed genomic and phenotypic analysis of four breast cancers through 2 to 15 years of diverse treatments, using a unique set of longitudinal samples from these patients. This revealed genetic events likely leading to drug resistance, including acquisition of *BRCA2* reversions and *ABCB1* fusions. Further, cancer phenotypes evolved dramatically after treatment, including increased post-treatment mesenchymal, receptor tyrosine kinase, and immune avoidance gene expression profiles. In one patient, treatment of cultured patient cells with drugs targeting the receptor tyrosine kinase phenotype was more effective in post-treatment, receptor tyrosine kinase-high cells compared to the pre-treatment cells. Thus, we have identified both mutations and phenotypes that may promote breast cancer drug resistance, some of which may be targeted to treat drug-resistant cancers.

TABLE OF CONTENTS

LIST OF FIGURES

Figures

ACKNOWLEDGEMENTS

Sincere thanks to my adviser, Andrea, for thinking big-picture and giving me the opportunity and training to work on a fantastic project. Thanks to Dr. John Hurdle, and to the NIH, for providing funding (NLM training grant 5T15LM007124) and introducing me to Andrea. I also thank Steve Piccolo for hands-on training in coding and bioinformatics, and the Marth and Quinlan labs for expert assistance and training. Dr. Theresa Werner, Dr. Saundra Buys, and Dr. Adam Cohen offered essential clinical help and aided in obtaining cancer specimens. Gaju Shrestha and Jasmine McQuerry provided excellent experimental help. Several patients generously donated their specimens to this research, without which this thesis would not have been possible. Finally, I thank my wife, Kaitlyn, and my four sons for being patient and sustaining me lovingly through this work.

CHAPTER 1

INTRODUCTION

Cancers arise from somatic mutations leading to unrestrained proliferation. They are genetically unstable due to harboring loss-of-function mutations in DNA repair pathways or other mechanisms[1]. This genomic instability leads to continuous evolution, even after the cancer has already been established. Each cell division, mutations occur; occasionally, these mutations lead to a selective advantage, and cells with the mutation out-compete neighboring cells until a cancer subclone (a single-cell-derived genetic lineage of cancer cells comprising a substantial proportion of cancer cells in the body) is established. Some mutations provide a selective advantage in the face of drug treatment, thus promoting acquired cancer drug resistance[2]. These mutations may or may not provide a selective advantage under normal conditions, but after treatment may come to dominate the cancer cell population as other non-resistant subclones die during initial response to the drug. A better understanding of which mutations promote drug resistance in cancer could lead to more effective therapeutic options.

The mutations promoting acquired drug resistance in breast cancer are imperfectly understood. One study found that few commonly occurring mutations occurred after treatment in a cohort of breast cancer patients who were studied both before and after

treatment[3]. Other studies have identified drug resistance-promoting mutations appearing in cultured breast cancer cells after long-term exposure to chemotherapy[4]. However, it is unclear whether these drug resistance mechanisms occur in actual patients, which motivated us to study pre- and post-treatment breast cancers directly from patients to identify possible mutations promoting drug resistance using DNA sequencing (DNA-Seq). Further, the few studies that have analyzed pre- and post-treatment breast cancers have explored each breast cancer's evolution after one treatment regimen[3], rather than tracking their sequential evolution through multiple regimens.

In addition to acquired mutations, drug-resistant cancers may acquire biological phenotypes promoting drug resistance[2]. These phenotypes may not be caused by any one obvious drug resistance mutation, but may nevertheless suggest therapeutic vulnerabilities. Thus, we sought to study both the genetic evolution (largely through DNA-Seq) and phenotypic evolution (largely through RNA-Seq) of breast cancers before and after multiple successive treatments. This led to the identification of both genetic events (Chapter 2) and phenotypes (Chapter 3) appearing post-treatment in breast cancers, which suggested treatment options, including one treatment option which we validated in a drug assay of cultured patient cells.

CHAPTER 2

GENETIC EVOLUTION OF FOUR BREAST CANCERS

Introduction

An understanding of the genetic evolution of breast cancers after treatment may suggest alternative treatment options for drug-resistant breast cancer. Thus, we studied the genetic evolution of four breast cancers over 2-15 years, during which all classes of breast cancer treatments were administered.

Materials and Methods

Overview

We performed DNA-Seq (whole-genome sequencing or WGS) on 2-6 longitudinal breast cancer samples from each of four patients, as well as their germline (blood-derived) samples. The acquisition of multiple longitudinal samples from the same patient, which is difficult and rare due to concerns and morbidities from repeated biopsy, was facilitated by using metastatic pleural effusion and ascites fluids from patients. Pleural effusions are fluid build-up in the pleura, a space surrounding the lungs, due to metastatic cancer cell infiltration[5]. Ascites are similar but occur in the peritoneal cavity[6]. Some 10-15% of breast cancer patients have pleural and/or ascitic metastasis[5], but these samples have been under-utilized for cancer research, other than for cell line generation[7].

In contrast to tumor biopsies, which hold inherent risks and discomforts, drainage of these cancer cell-rich chest and abdominal fluids is relatively safe and desirable for the patient, for whom the fluid buildup can be extremely uncomfortable[5]. Some patients required frequent, even weekly, drainage, enabling the tracking of their cancer's evolution with good resolution.

In addition to identifying potential drug resistance mutations, WGS can help with identification of the general subclone evolution of each patient's cancer. This is because each subclone tends to have dozens, if not hundreds, of mutations that evolve in a similar direction as a tell-tale indicator of that subclone's evolution. For example, a group of mutations appearing after treatment with a modal variant allele frequency (VAF; percent of sequencing reads with the mutation) of 0.5 indicates a new, completely dominant subclone representing 100% of the cells (since most mutations are heterozygous, a VAF of 0.5 generally indicates a heterozygous present in all cells). On the other hand, a group of mutations appearing after treatment with a modal VAF of 0.3 indicates a new subclone representing 60% cellular prevalence. DNA-Seq of individual cells can be used to verify findings thus inferred. Below, I describe detailed methodology for isolating cancer samples, performing DNA-Seq, determining mutational subclone evolution, and performing related experiments on samples from our four breast cancer patients.

Samples and processing

Breast cancer samples were obtained from malignant pleural effusions or ascites (or FFPE in one case). After fluid drainage, cells were pelleted at 3,750 RPM for 5 minutes. Cells were resuspended in buffer (17 mM Tris, pH 7.4, and 135 mM NH_4Cl)

and incubated at 37 °C for 5 minutes, followed by centrifugation, which was repeated until red blood cells were absent from pellet. Cells were washed thrice with PBS and frozen in 90% FBS with 10% DMSO. The patient #2 day 1320 pleural effusion had cancerous chunks from which DNA or RNA was isolated directly. For scRNA-Seq of this sample, minced chunks were dissociated in Renaissance medium (Cellaria) with 1% TrypLE (Life Technologies) and 1 U/μL DNase I and 2 mg/mL collagenase (Roche) for 20 minutes at 37 °C (shaking occasionally), followed by 10 minutes in a Biomaster 80 (Seward) and passing chunks through a 21-gauge syringe, and freezing in 10% DMSO for storage until scRNA-Seq.

DNA isolation and sequencing

Frozen vials of pleural effusion or ascites cells were thawed, and cells were pelleted to remove DMSO in supernatant and re-suspended in buffer as described in Miltenyi OctoMACS and QuadroMACS instructions. CD45+ white blood cells were depleted from non-tumor-chunk samples (most samples) using the Miltenyi OctoMACS magnetic separation system with MS columns or the QuadroMACS magnetic system with LD columns along with anti-CD45 microbeads from Miltenyi. Germline DNA was obtained from PBMCs from each patient or by sorting for CD45+ white blood cells from pleural effusions. DNA was isolated using Qiagen's QIAamp DNA Micro Kit, or for FFPE, the QIAamp DNA FFPE Tissue Kit. Whole-exome sequencing (WES) was performed at the Huntsman Cancer Institute's High Throughput Genomics Core Facility (HCI) using the Agilent SureSelect QXT Human All Exon v5 + UTRs kit and library prep and an Illumina HiSeq 2500 instrument with 125 cycles and paired-end sequencing.

Whole-genome sequencing (WGS) was performed at the McDonnell Genome Institute at Washington University, NantOmics, or HCI using the Illumina TruSeq Nano DNA Library Prep Kit or PCR-free library prep and an Illumina HiSeq 2500 or X instrument with paired-end sequencing.

Single-cell DNA sequencing

Frozen viable pleural effusion or ascites vials were thawed and CD45+ white blood cells were depleted using Miltenyi's QuadroMACS. Individual cells were captured using Fluidigm C1 chips (10-17 μm cell diameter); whole-genome amplification was performed per manufacturer's instructions. Targeted amplification of mutation-containing regions was performed using Fluidigm Access Array and BioMark instruments per manufacturer's instructions. Libraries were sequenced on Illumina MiSeq. Reads were aligned to hg19 with BWA MEM v0.7.8. Variants were called using MuTect v1.1.4. Co-occurrence of mutations (at least 1 mutant read) in individual cells was evaluated using Fisher's exact test. Single cells with more than one subclone-defining mutation are shown.

Whole-exome sequencing variant identification

Read quality was verified using FastQC and reads were trimmed using Trimmomatic[8] v0.32 (http://www.usadellab.org/cms/index.php?page=trimmomatic). Alignment to hg19 was done with BWA MEM[9] v0.7.8 (https://github.com/lh3/bwa). BAM files were refined using PicardTools' *MarkDuplicates* and *FixMateInformation* tools, BamTools[10] *filter* tool (https://github.com/pezmaster31/bamtools), and GATK's[11]

RealignerTargetCreator, *IndelRealigner*, and *BaseRecalibrator* (v3.2-2; https://www.broadinstitute.org/gatk/download/). WES somatic SNVs were called with MuTect[12] v1.1.4 (https://www.broadinstitute.org/cancer/cga/mutect) and annotated with Oncotator[13] v1.3.0.0 (https://github.com/broadinstitute/oncotator). CNVs were called from WES using VarScan[14] v2.3.7 (http://dkoboldt.github.io/varscan) *copynumber* and *copyCaller*, from pileups generated by SAMTools, to determine copy-number 2 genes for SubcloneSeeker. Segmentation was done with DNAcopy in R. Segmented data were converted to gene level using UCSC's refGene.txt (hg19) annotation. Log_2 fold-change values were converted to absolute copy using 2^{n+1} where n is log_2 fold-change from diploid. CNV values were shifted in each sample to make the 2-copy peak centered at 2 and multiplied around the 2-axis to maximize number of genes near (within 0.1) of 1, 3, and 4 (to adjust for normal contamination).

Whole-genome sequencing variant identification

WGS DNA sequences were aligned to hg19 using BWA MEM using SpeedSeq[15]. SNVs and indels variants were identified using FreeBayes (http://arxiv.org/abs/1207.3907) on each of the four breast cancer patients analyzed. Variants were annotated using SnpEff[16] and variants with quality below 5 were excluded. Somatic mutations were identified using "somatics" (https://github.com/brentp/gobio/tree/master/somatics). As a secondary variant calling approach, we used VarScan *somatic*. This revealed the *APC* and *BRCA2* K817* mutations in patient #4 not detected by FreeBayes; all other small variants discussed were identified with FreeBayes.

Structural variants were detected from WGS as follows. We used SAMBLASTER[17] v0.1.22 to extract discordant paired-end reads and split reads. LUMPY[18] v0.2.12, or LUMPY within the SpeedSeq suite (*sv* utility), was used to call structural variants based on these reads, followed by SVTYPER[15] (v0.0.2) to determine VAFs for each variant. Somatic variants were identified as with FreeBayes variants. Variants were annotated with BEDTOOLS[19] v2.25.0-12-g7f765bc. Variants with quality less than 400 were excluded, except for patient #2 *ABCB1* fusions, which did not meet quality thresholds but were included due to corroborating identification in RNA-Seq data. Circos[20] plots were made from FreeBayes, LUMPY, and CNV analysis. Circos plots show evolving mutations (going from VAF below 0.05 to 0.05 or above for small mutations and a threshold of 0.075 for structural) in color and background variants in grey; Cancer Gene Census[21] genes are shown by name, while other mutations are indicated by tick marks. Copy number variation (CNV) was determined from WGS data using VarScan as described for WES. For Circos plots CNV tracks, 30-segment window averages (absolute copy) were plotted after adjustment for normal contamination.

Correction for normal contamination

Variant allele frequencies (VAFs) for WGS FreeBayes-identified somatic variants were corrected for normal contamination by determining tumor purity using CNV data. Absolute copy of each gene was determined using VarScan and DNAcopy in R as described in previous sections. Perfectly pure tumor samples have a large copy peak at 2 and smaller peaks at 1, 3, and 4, while normal-contaminated samples have a profile collapsing towards 2. We centered the largest peak at 2 and then multiplied values around

the 2-axis using a range of multipliers until the maximum number of genes possible fell within 0.1 of 1, 3, and 4. The multiplier thus obtained was used to calculate tumor purity (= 1.0 / multiplier), and VAFs for FreeBayes somatic mutations were multiplied by the sample's multiplier to obtain adjusted VAF. VAFs used and reported are all adjusted except for the germline *BRCA2* germline V643fs mutation VAFs in patient #4 given its presence in both normal and cancer cells.

Identification of evolution clusters and subclones from WGS data

Clusters of co-evolving mutations were identified from the WGS SNV and indel data (after normal contamination correction) by clustering copy-neutral (absolute copy 1.5-2.5 or, in the case of pseudo-tetraploid patient #2, between 3.6-4.4) mutations based on the samples in which they were present (at variant allele frequency of at least 0.05). Consensus VAFs for each mutation cluster were determined using kernel density estimation in R; cellular prevalences for each cluster were then obtained by multiplying these values by 2 (or 4 for pseudo-tetraploid patient #2). In some cases, two peaks were present, indicating different sub-clusters and possibly additional subclones. Mutation clusters with the most mutations were incorporated into subclone analysis. Subclone structures were then determined using cellular prevalences thus calculated, using rules described elsewhere[22]. For patient #1, an additional subclone was made apparent from deeper WES and single-cell DNA sequencing (purple mutations) that was not detectable by WGS.

Identification of subclones using SubcloneSeeker

We started with diploid, missense SNVs identified from deep WES by MuTect for patient #1 to corroborate WGS-based subclone findings. These variants were then subjected to Affinity Propagation Clustering[23] (via R package apclust (https://cran.r-project.org/web/packages/apcluster/citation.html)), with similarity calculated via expSimMat(r=2, w=0.1), to identify groups of variants that share similar VAFs across all samples. We then performed subclone structure reconstructions at adjacent pairwise time-points through the pairwise joint analysis capability of SubcloneSeeker[22], and manually merged structures into a longitudinal evolution history, abiding to the same evolutionary consistency rule implemented in pairwise merging.

Mutational signatures

Mutational signatures were identified for WGS SNVs by first determining the trinucleotide sequence around SNV mutation sites, with the mutation site in the center, using an in-house script. For trinucleotides starting with an A or G, the reverse complement was used to minimize the number of mutation contexts as reported previously[24]. From these data, we identified the percent of mutations that fell into each trinucleotide and transition/transversion pattern.

TopHat Fusion

TopHat[25] v2.0.6 (http://ccb.jhu.edu/software/tophat/index.shtml) was run on bulk RNA-Seq data using hg19 followed by TopHat-Fusion Post[26]. This identified the *SLC25A40-ABCB1* fusion transcript in patient #1 but missed the *ABCB1* fusions in

patient #2, which we identified manually by searching for reads fusing the 3' end of *ABCB1* exon 2 with transcripts aligning elsewhere (Blast).

qRT-PCR

HCC70 cells grown in DMEM/F12 with 9% FBS were treated for 24 hours with indicated doses of paclitaxel, followed by RNA isolation with RNAzol, reverse transcription with SuperScript III Reverse Transcriptase (Life Technologies), and qPCR using KAPA SYBR FAST with primers to *APOBEC3B* or *ACTB*. Data were analyzed using ΔΔCt to obtain *ACTB*-normalized *APOBEC3B* mRNA expression.

DNA content measurement

Frozen viable pleural effusions were thawed and washed. Cells were resuspended in 4% paraformaldehyde in PBS and fixed 15 minutes. Cells were pelleted and resuspended in 5 mL PBS for 15 minutes. Cells were pelleted and resuspended in staining buffer (2% FBS in PBS). 3 µM DAPI was added and incubated 15 minutes. Cells were washed twice in PBS and resuspended in PBS, followed by flow cytometry.

Results and Discussion

We used a systems approach to determine the genetic and phenotypic evolution of four metastatic ER+ breast cancers over 2 to 15 years and 3 to 6 samples per patient. For each patient, we determined subclonal evolution through bulk and/or single-cell DNA sequencing at various points in the patient's treatment history (Fig. 1, #1 and #2). Further, we used RNA-seq to identify biological phenotypes associated with these evolving

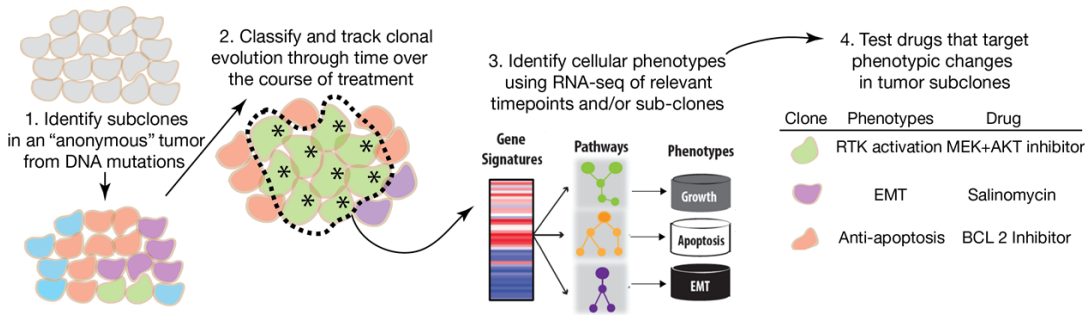


Figure 1 | Overview of systems approach for identifying therapeutic vulnerabilities from longitudinal genomic analysis. * = resistant subclone. Made in conjunction with and used by permission of Aaron Quinlan and Evan Johnson.

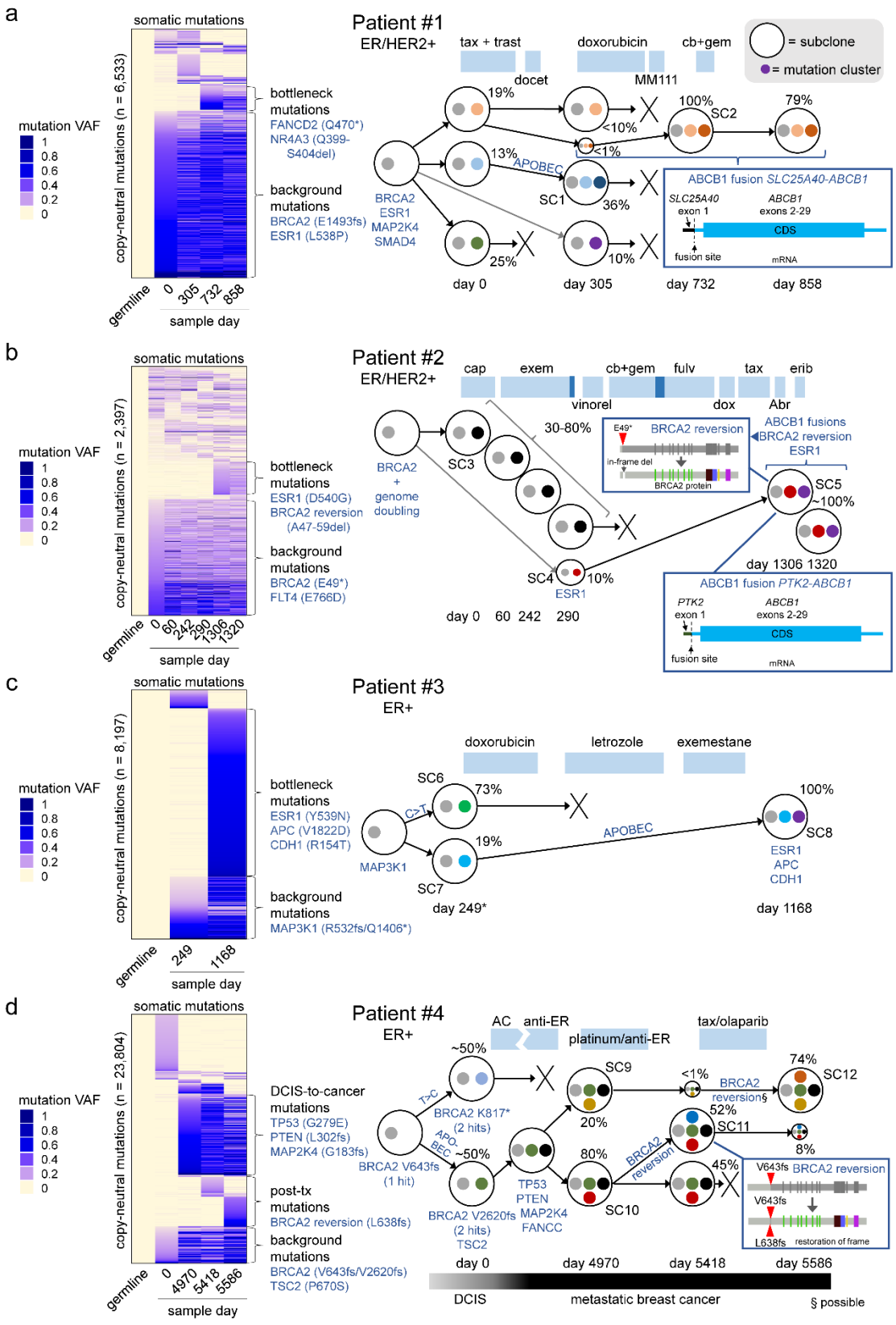
subclones, and to identify drugs predicted to target them (covered in Chapter 3). This identified effective treatments for post-chemotherapy subclones, as shown by drug assays using patient tumor cells (Fig. 1; #3 and #4).

To determine the subclonal evolution of our four breast cancers, we used 60× whole-genome (WGS), 100× whole-exome (WES), and targeted single-cell DNA sequencing, along with the SubcloneSeeker[22] tool. First, somatic WGS mutations found in copy-neutral regions were grouped into mutation clusters based on their presence or absence at specific timepoints. Next, the cellular prevalence of each mutation cluster was determined by identifying the variant allele frequency of maximum density at each timepoint, and subclones were then identified using rules previously described[22]. For patient #1, this approach was validated by SubcloneSeeker analysis of WES data (Supplementary Fig. 1) and refined by single-cell DNA or RNA sequencing, which confirmed SubcloneSeeker predictions and refined subclone identification (Supplementary Fig. 2). (See Materials and Methods.) Subclones are represented as circles containing mutation clusters (smaller colored circles within subclones) in Fig. 2, with cellular prevalence of each subclone indicated as a percentage next to each subclone. In addition, variant allele frequencies for copy-neutral mutations (SNVs and indels) used to determine subclone evolution are shown as a heatmap to the left of each subclone schematic in Fig. 2. Copy-number and structural variants for each patient are shown in Circos plots in Supplementary Fig. 3. Mutations in "cancer genes" according to the Cancer Gene Census[27] are shown by name in Fig. 2 and in Supplementary Fig. 3.

Patient #1 (ER+/HER2+) samples were collected over more than 2 years from malignant pleural effusion and ascites fluids. Samples were obtained before, during, or

Figure 2 | Subclonal evolution of four breast cancers over 2-4 years. (**a-d**) Subclonal evolution of breast cancer patients #1 to #4 through treatment. Left side shows variant allele frequencies of copy-neutral somatic mutations (WGS) organized into clusters, with relevant cancer-associated mutations (may or may not be copy-neutral or somatic) indicated. Right shows subclone evolution. Subclones are indicated by large circles; mutation clusters are indicated by small colored circles. Relevant mutations in subclones are indicated by text or boxed insets. Cellular prevalence is indicated as percent next to subclone. Tax, paclitaxel; trast, trastuzumab; dox, doxorubicin; cb, carboplatin; gem, gemcitabine; cap, capecitabine; exem, exemestane; vinorel, vinorelbine; fulv, fulvestrant; Abr, Abraxane; erib, eribulin. *Patient #3 day 0 had RNA-Seq only.

Figure 2


a
somatic mutations
mutation VAF
1
0.8
0.6
0.4
0.2
0
copy-neutral mutations (n = 6,533)
bottleneck mutations
FANCD2 (Q470*)
NR4A3 (Q399-S404del)
background mutations
BRCA2 (E1493fs)
ESR1 (L538P)
germline
0
305
732
858
sample day
Patient #1
ER/HER2+
tax + trast
docet
doxorubicin
cb+gem
MM111
= subclone
= mutation cluster
19%
<10%
100%
SC2
79%
13%
APOBEC
<1%
SC1
36%
BRCA2
ESR1
MAP2K4
SMAD4
25%
10%
ABCB1 fusion SLC25A40-ABCB1
SLC25A40
exon 1
ABCB1
exons 2-29
CDS
fusion site
mRNA
day 0
day 305
day 732
day 858
b
somatic mutations
mutation VAF
1
0.8
0.6
0.4
0.2
0
copy-neutral mutations (n = 2,397)
bottleneck mutations
ESR1 (D540G)
BRCA2 reversion (A47-59del)
background mutations
BRCA2 (E49*)
FLT4 (E766D)
germline
0
60
242
290
1306
1320
sample day
Patient #2
ER/HER2+
cap
exem
cb+gem
fulv
tax
erib
vinorel
dox
Abr
30-80%
E49*
BRCA2 reversion
in-frame del
BRCA2 protein
ABCB1 fusions
BRCA2 reversion
ESR1
BRCA2
+
genome
doubling
SC3
SC5
~100%
day 1306 1320
SC4
10%
ESR1
ABCB1 fusion PTK2-ABCB1
PTK2
exon 1
ABCB1
exons 2-29
CDS
fusion site
mRNA
day 0
60
242
290
c
somatic mutations
mutation VAF
1
0.8
0.6
0.4
0.2
0
copy-neutral mutations (n = 8,197)
bottleneck mutations
ESR1 (Y539N)
APC (V1822D)
CDH1 (R154T)
background mutations
MAP3K1 (R532fs/Q1406*)
germline
249
1168
sample day
Patient #3
ER+
doxorubicin
letrozole
exemestane
SC6
73%
C>T
19%
APOBEC
100%
SC8
MAP3K1
SC7
ESR1
APC
CDH1
day 249*
day 1168
d
somatic mutations
mutation VAF
1
0.8
0.6
0.4
0.2
0
copy-neutral mutations (n = 23,804)
DCIS-to-cancer mutations
TP53 (G279E)
PTEN (L302fs)
MAP2K4 (G183fs)
post-tx mutations
BRCA2 reversion (L638fs)
background mutations
BRCA2 (V643fs/V2620fs)
TSC2 (P670S)
germline
4970
5418
5586
sample day
Patient #4
ER+
AC
anti-ER
platinum/anti-ER
tax/olaparib
~50%
SC9
<1%
BRCA2 reversion§
74%
SC12
T>C
BRCA2 K817*
(2 hits)
20%
52%
SC11
APO-BEC
~50%
80%
8%
BRCA2 V643fs
(1 hit)
TP53
PTEN
MAP2K4
FANCC
BRCA2 reversion
45%
V643fs
BRCA2 reversion
BRCA2 V2620fs
(2 hits)
TSC2
SC10
V643fs
L638fs
restoration of frame
day 0
day 4970
day 5418
day 5586
DCIS
metastatic breast cancer
§ possible

after the following treatments: (1) paclitaxel and trastuzumab followed by docetaxel, (2) liposomal doxorubicin, (3) trastuzumab and MM-111 (an experimental HER2/HER3 antagonist[28]) followed by (4) carboplatin and gemcitabine (Fig. 2a), after which the patient succumbed to disease. Background (ubiquitous) mutations in patient #1 included a germline *BRCA2* E1493fs mutation that underwent loss of heterozygosity in her cancer, an *ESR1* L538P (likely activating[29]) mutation, and large homozygous structural variants (likely inactivating) in *SMAD4* and *MAP2K4* (Fig. 2a, Supplementary Fig. 3a). Following a response to paclitaxel with trastuzumab, the patient acquired three new subclones, suggesting independent acquired resistance mechanisms to this treatment ("tax + trast"; see Fig. 2a). One of these subclones, SC2, appeared at a low cellular prevalence after paclitaxel and trastuzumab (<1%) but came to dominate at 100% cellular prevalence after subsequent treatment with liposomal doxorubicin ("doxorubicin" in Fig. 2a), to which the patient partially responded. This bottleneck subclone possessed an *SLC25A40-ABCB1* fusion resulting from a structural deletion on chromosome 7, which fused the active promoter and 5' UTR of *SLC25A40* to the *ABCB1* gene, preserving the entire *ABCB1* coding region and leading to increased levels of the ABCB1[4] drug efflux pump (Fig. 2a, inset and Supplementary Figs. 4, 5a). SC2 likewise possessed a copy gain on chromosome 10, including *CDK1*, and copy loss on chromosome 2 (Supplementary Fig. 3a, blue arrows). In addition, the bottleneck subclone SC2 possessed 1,047 new SNVs and indels ("bottleneck mutations" in Fig. 2a, left; dark-orange mutation circles found in Fig. 2a, right), including mutations in the cancer-associated genes *NR4A3*[30] and *FANCD2*[31].

Interestingly, patient #1 also possessed a subclone that acquired numerous

APOBEC-associated[24] mutations after paclitaxel and trastuzumab (subclone SC1; Fig. 2a, Supplementary Fig. 6, third column), with 31.3% of subclone-specific SNVs in this subclone being APOBEC-associated compared to 10.0% of subclone-specific variants in its parental subclone (light-blue mutations in Fig. 2a), and 12.7% of background (grey) variants (Supplementary Fig. 6, first and second columns). This mutational process may have been promoted by the patient's taxane treatment as paclitaxel increased *APOBEC3B* expression *in vitro* (Supplementary Fig. 7). Thus mutational processes may differ between subclones within a tumor and may be promoted by certain treatments.

Patient #2 (ER+/HER2+) was followed for over 3 years, including 6 pleural effusions. Samples were obtained before, during, or after the following treatments: (1) capecitabine, (2) exemestane, (3) exemestane and everolimus, (4) vinorelbine, (5) carboplatin and gemcitabine, (6) fulvestrant, (7) liposomal doxorubicin, (8) paclitaxel, (9) Abraxane, and (10) eribulin (Fig. 2b). This patient was pseudo-tetraploid (Supplementary Fig. 8) due to an early genome doubling event, necessitating some modifications to subclone identification (see Materials and Methods) and making it unclear whether subclone SC4 was derived from SC3 and SC3's prevalence (Fig. 2b). Patient #2 possessed a germline heterozygous *BRCA2* E49* mutation that underwent loss of heterozygosity in the patient's cancer, and a possibly oncogenic *FLT4* E766D somatic mutation, at all timepoints. This patient initially possessed one major subclone (SC3) with less than 80% cellular prevalence (Fig. 2b). Towards the end of treatment with the anti-estrogen exemestane, a new minor subclone (SC4), with 10% cellular prevalence, appeared possessing an *ESR1* D540G mutation (Fig. 2b, Supplementary Fig. 3b), which likely promoted resistance to exemestane[29]. The patient's disease then began to progress,

requiring treatment with a wide variety of therapies. After more than 2 years and numerous treatment regimens, patient #2's cancer was dominated by a new subclone (SC5), derived from the *ESR1*-mutant subclone SC4, at 100% cellular prevalence (Fig. 2b). The new bottleneck subclone SC5 possessed 936 new SNVs and indels ("bottleneck mutations" in Fig. 2b heatmap). One of these, an in-frame *BRCA2* A47-P59 deletion, removed the inactivating frameshift in one copy of *BRCA2* and likely restored its function, as a resistance mechanism to platinum therapy[32] the patient had received (Fig. 2b, "BRCA2 reversion" inset, and Supplementary Fig. 9a). Additionally, SC5 acquired two *ABCB1* fusions (*PTK2-ABCB1* and *AFF3-ABCB1*) that, like the *SLC25A40-ABCB1* fusion, each provided *ABCB1* with a strong promoter while maintaining the entire *ABCB1* coding region intact (Fig, 2b, bottom-right inset and Supplementary Fig. 10), and apparently promoted *ABCB1* expression (Supplementary Fig. 5b). These fusions may have promoted resistance to doxorubicin and paclitaxel. It is unclear whether the two fusions were acquired sequentially due to selective pressure for additional *ABCB1* expression, or co-existed in the original refractory subclone. Subclone SC5 also lost a large region of chromosome 20q, including the tumor suppressor *GNAS* (Supplementary Fig. 3b, orange arrow). Unlike patient #1, this patient's mutational signatures, such as the APOBEC signature, were relatively uniform between subclones (Supplementary Fig. 6).

Patient #3 (ER+/HER2-) was followed for more than 3 years, including 3 pleural effusions. Samples were obtained before, during, or after: (1) two doxorubicin courses, (2) letrozole, and (3) exemestane (Fig. 2c). The patient showed clinical benefit from each of these treatments followed by progression. Background mutations included two *MAP3K1* mutations (R532fs and Q1406*) inactivating this tumor suppressor[33]. Initially,

two subclones existed: SC6, enriched in C>T mutations, dominated at 73% cellular prevalence, while subclone SC7 represented 19% of cells (Fig. 2c, Supplementary Fig. 6). However, after doxorubicin, letrozole, and exemestane treatment, a new SC7-derived subclone came to dominate at 100% cellular prevalence (SC8; Fig. 2c, Supplementary Fig. 3c). This bottleneck subclone possessed a striking 5,540 additional SNVs and indels (Fig. 2c, "bottleneck mutations"), nearly three times the background mutation number (n = 2,042). Of the new SNVs, 67.4% were APOBEC-associated, compared to 27.5% of background (grey) SNVs (Fig. 2c, Supplementary Fig. 6). Among the 5,540 new mutations in SC8 were *ESR1* Y539N, which likely promoted resistance to letrozole or exemestane anti-estrogen treatments[29], *APC* V1822D, and *CDH1* (E-cadherin) R154T mutations (Fig. 2c, Supplementary Fig. 3c). Additionally, the bottleneck subclone SC8 gained copies of chromosome 5p, and lost regions of chromosome 1p, a tumor-suppressive region[34,35], 2q, 15, and 22 (Supplementary Fig. 3c, green arrows). Lost regions included the apoptosis genes *CASP8*, *CASP10*, and *BID*[36], loss of which could prevent apoptosis.

Patient #4 (ER+/HER2-) was followed for 15 years, including her primary tumor (ductal carcinoma *in situ* or DCIS) and recurrent pleural effusions, including the following treatments: (1) doxorubicin with cyclophosphamide ("AC" in Fig. 2d), (2) the anti-estrogens tamoxifen and letrozole, with a period of capecitabine between, (3) cisplatin followed by fulvestrant, interferon treatment for a non-cancer diagnosis, carboplatin, then tamoxifen ("platinum/anti-ER"), and (4) paclitaxel followed by olaparib ("tax/olaparib"; Fig. 2d). Background mutations in patient #4 included a *BRCA2* V643fs germline mutation and a *TSC2* P670S mutation (Fig. 2d). The day 0 DCIS possessed two

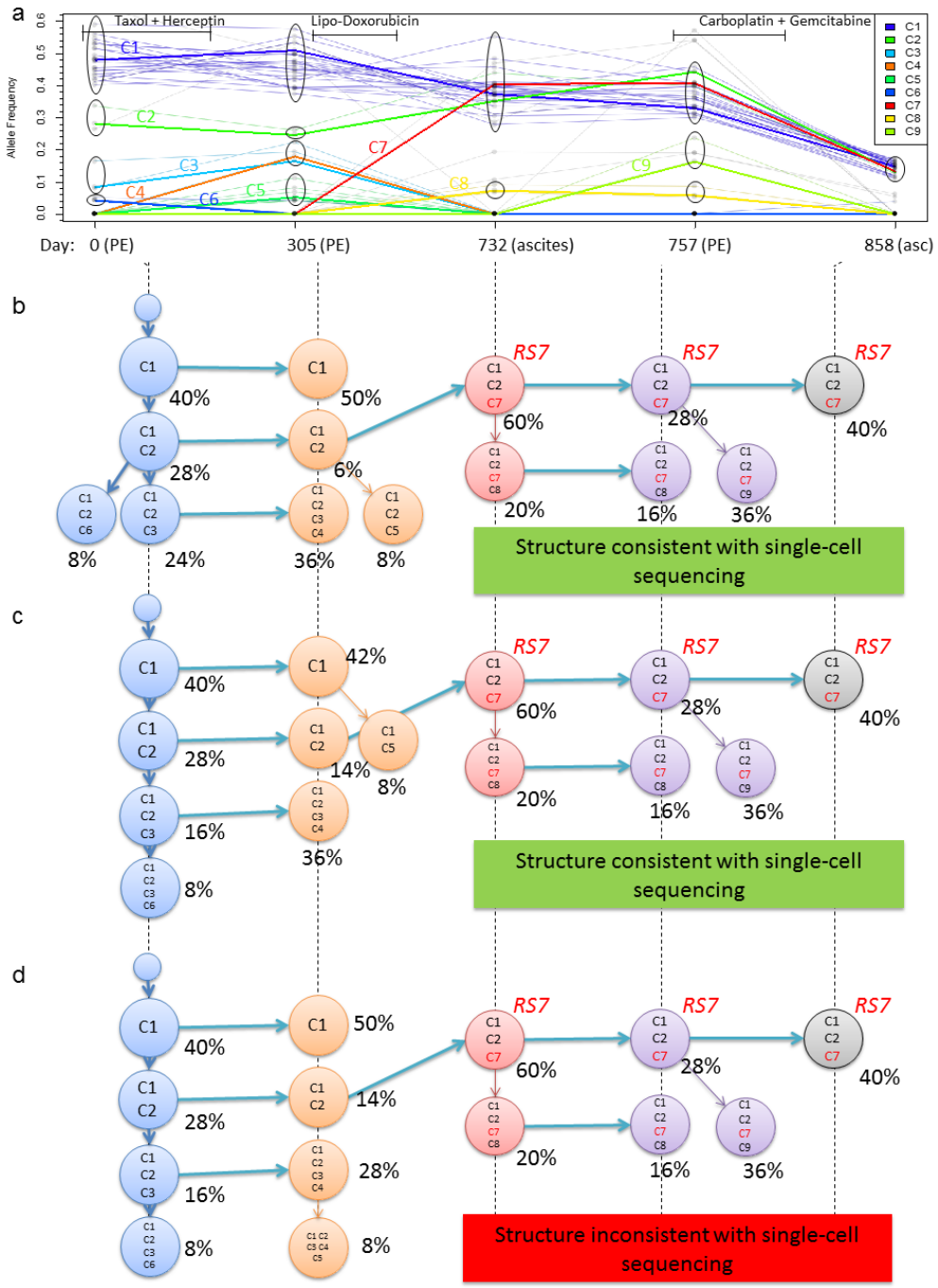
independent cancers at ~50% cellular prevalence each, with independent *BRCA2* second-hit mutations: a *BRCA2* K817* mutation, which subsequently disappeared; and a *BRCA2* V2620fs mutation, representing the surviving subclone giving rise to subsequent samples (Fig. 2d, Supplementary Fig. 3d). Thirteen years later, after transition from DCIS to metastatic disease, the patient had acquired additional mutations in *TP53* (G279E homozygous), *PTEN* (L302fs and N329fs), *MAP2K4* (G183fs homozygous), and *TGFBR2* (G68S); these may have promoted transition from DCIS to metastatic cancer. At this first metastatic timepoint (day 4970), there were two major subclones, SC9 (20% cellular prevalence) and SC10 (80%). Subsequently, the patient received platinum-based therapy (cisplatin and carboplatin), which selected for a subclone of SC10, SC11, with a *BRCA2* reversion (L638fs) restoring the frame of the V643fs mutation after only a few aberrant amino acids, likely restoring BRCA2 function (Fig. 2d, bottom-right inset and Supplementary Fig. 9b). After further treatment with paclitaxel and olaparib, the BRCA2-revertant subclone SC11 decreased from 52% to 8% cellular prevalence, while SC12 became dominant at 74%. *BRCA2* reversions are thought to also promote olaparib resistance[37]; thus, it is unclear why SC11 decreased in prevalence after olaparib. One potential explanation is that subclone SC12, which survived olaparib, may have acquired a direct reversion of the germline *BRCA2* V643fs mutation, as the V643fs variant allele frequency went down to 0.28 at day 5586 from an average of 0.51 in all other samples, including germline (Supplementary Fig. 11). In addition, patient #4's DCIS-to-metastatic conversion was accompanied by enrichment in APOBEC-associated mutations, while SC12 was also enriched in APOBEC mutations, and SC11 had enrichment in a pair of C>T mutation types (Supplementary Fig. 6).

While each patient's subclone evolution differed, several themes emerged. First, effective treatments resulting in long-term cancer control generally resulted in genetic bottleneck events in which one major subclone survived treatment (subclones SC2, SC5, and SC8). Second, two patients acquired *ABCB1* promoter fusions previously unreported in breast cancer, likely promoting drug resistance. Third, *BRCA2* reversion events may be common in platinum-treated *BRCA2* carriers, and can occur through multiple mechanisms. Lastly, mutational signatures substantially differed between subclones, which may affect clonal fitness. Such mutational signatures can apparently be acquired *de novo*, independent of the parent subclone's mutational processes, and may be drug-induced (Supplementary Figs. 6, 7).

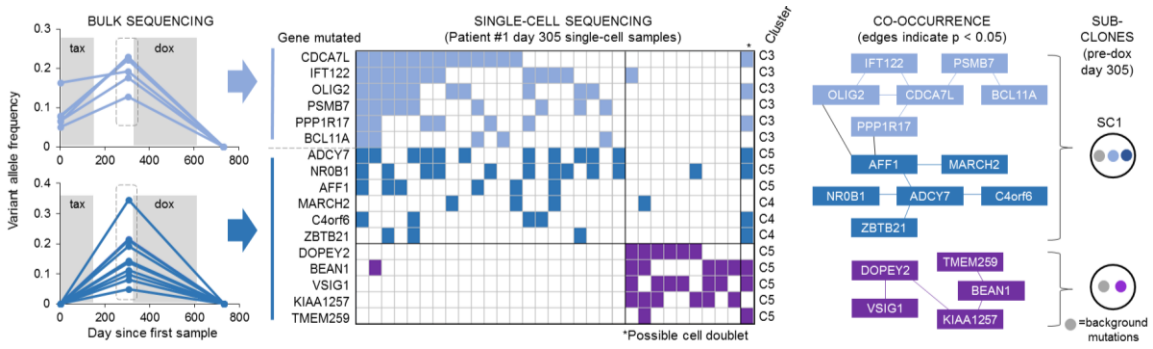
These data suggest that targeting *ABCB1* and *BRCA2* reversions, perhaps in combination with standard chemotherapy, may overcome drug resistance. Targeting the ABCB1 protein directly has led to toxicities due to ABCB1's essential function in the blood-brain barrier[38] and protection of stem cells[39]. However, one alternative approach might be to target the fused promoter driving *ABCB1* mRNA expression by inhibiting transcription factors essential for transcription initiation from that promoter. This could be done by identifying small molecules that inhibit the expression of the native gene of the promoter (*SLC25A40*, *PTK2*, or *AFF3*) using an approach such as the Connectivity Map[40]. This approach could lead to less toxicity if inhibiting expression of the fusion partner is less toxic than inhibiting ABCB1 itself, and merits further study.

Supplemental Material

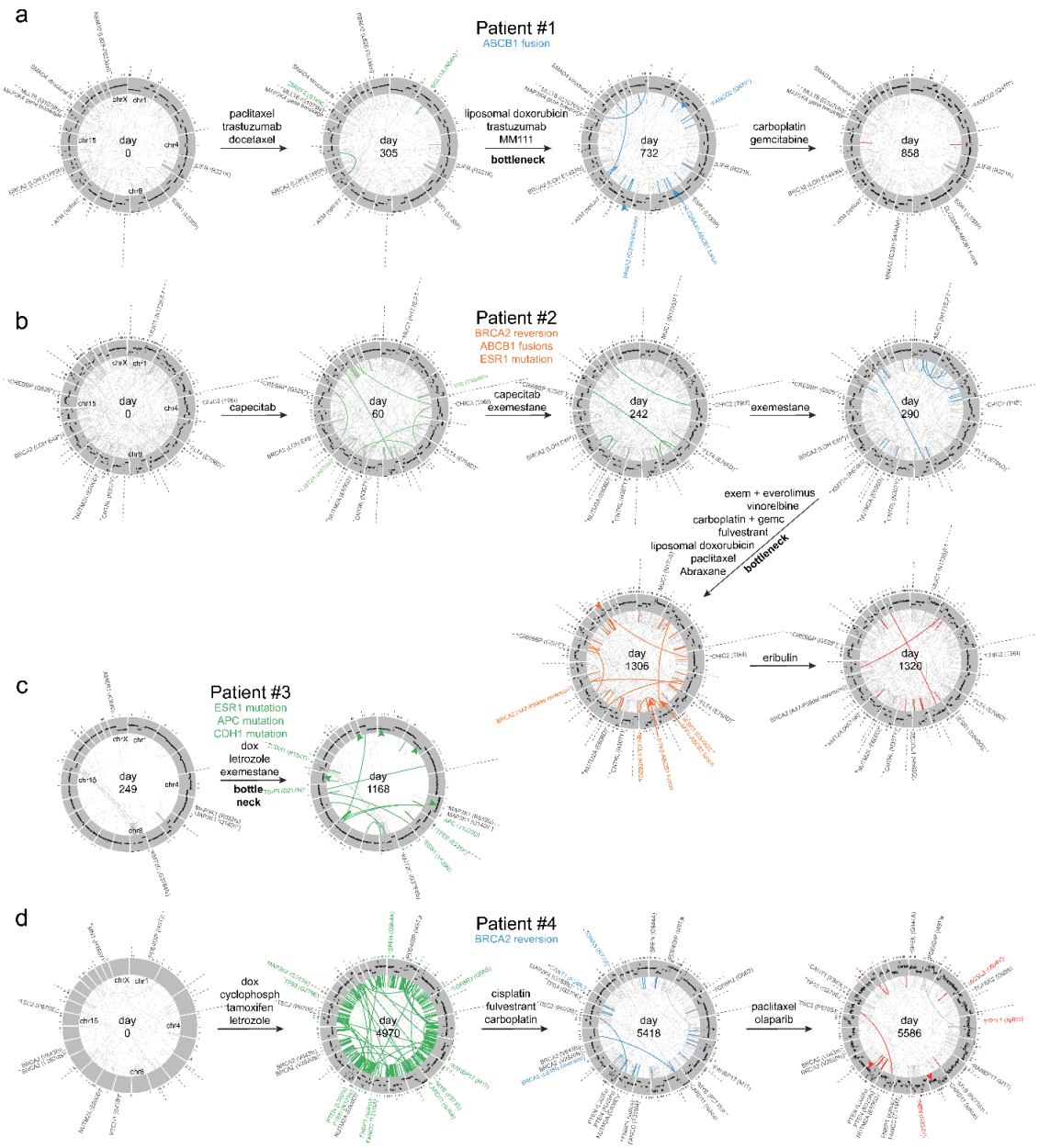
Supplementary figures referred to in this chapter can be found on the ensuing pages.



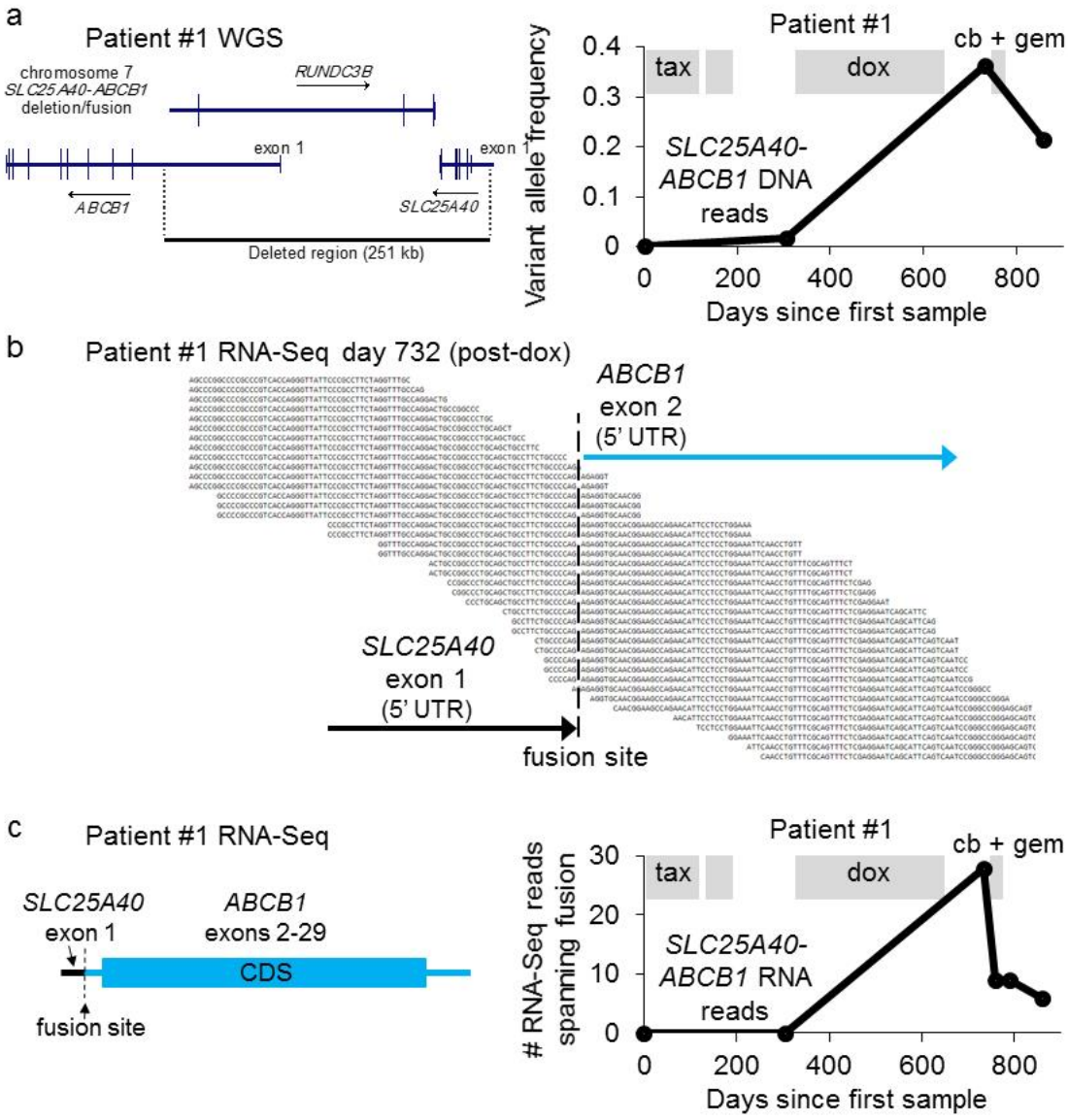
Supplementary Figure 1 | Possible subclonal evolution structures identified by SubcloneSeeker for breast cancer patient #1. (**a**) Variant allele frequencies for coding, diploid mutations identified from exome sequencing. Mutation clusters (C1-C9), identified by Affinity Propagation Clustering, are indicated by colors. (**b-d**) Possible subclonal evolution structures identified through analysis of mutations in (a) with SubcloneSeeker. Cellular prevalences are indicated next to each subclone. (b) and (c) were found to be consistent with single-cell sequencing while (d) was not. RS7 indicates resistant (bottleneck) subclone from cluster C7. Used with permission of Yi Qiao, Gabor Marth, and Stephen Piccolo.



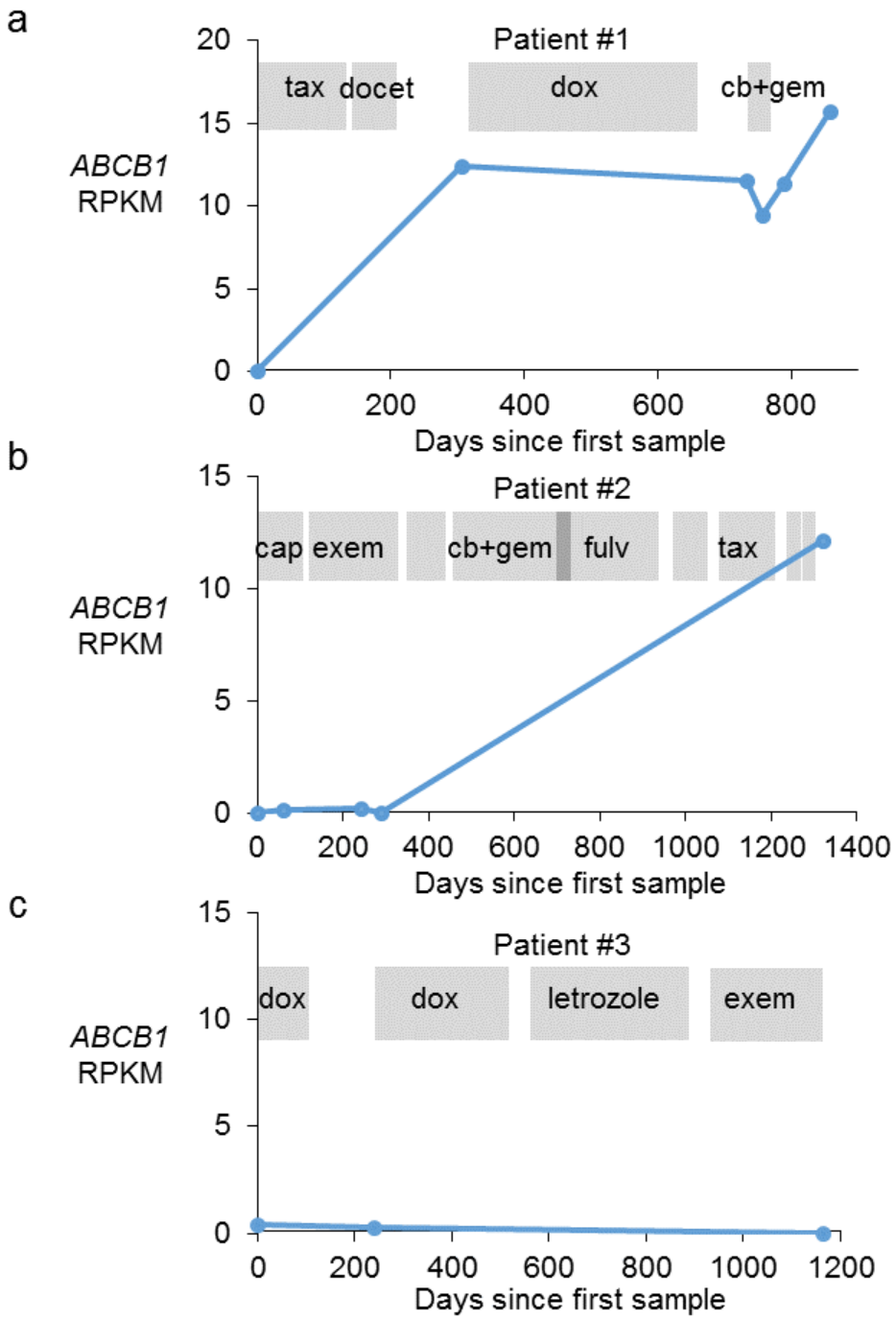
Supplementary Figure 2 | Single-cell sequencing of pre-doxorubicin cells in breast cancer patient #1. Left, bulk variant allele frequencies of mutations increasing after taxane therapy studied by single-cell sequencing as determined by bulk whole-exome sequencing (WES). Middle, results of single-cell sequencing of pre-dox cells (day 305), with each column representing a single cell, mutations sequenced indicated by the name of the gene (or nearby gene) mutated, and mutations indicated by filled squares; mutation clusters identified in Supplementary Fig. 1 are also indicated. Right, co-occurrence analysis of mutations in single cells; edges indicated $p < 0.05$ by Fisher's exact test. Subclones from Fig. 2a are indicated for the groups of mutations.



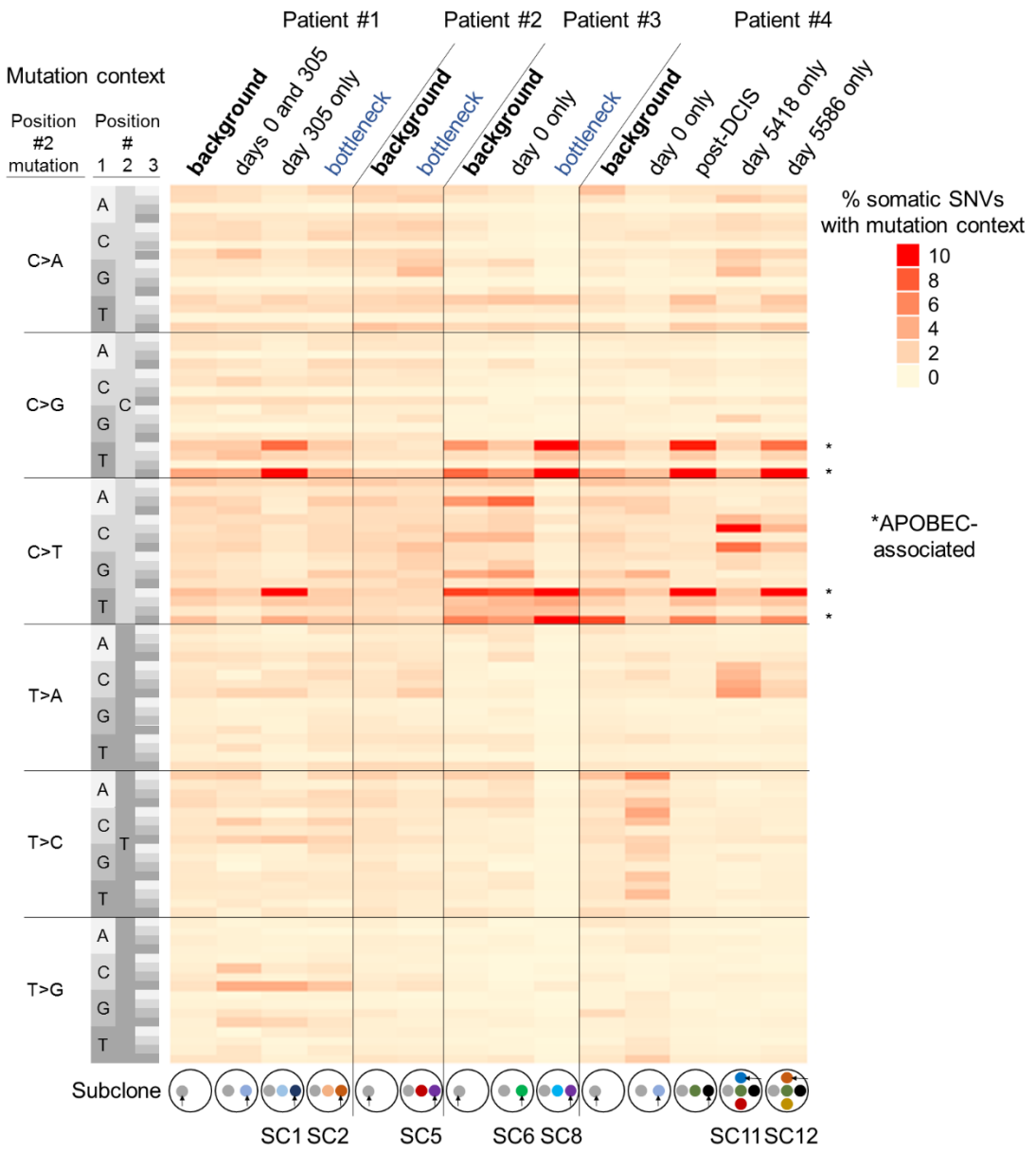
Supplementary Figure 3 | SNV, structural variant, and CNV evolution in four breast cancers. (**a-d**) Circos plots showing mutation evolution of breast cancer patients #1 to #4 through treatment. Each circle represents chromosomes 1 through X (clockwise from top) arranged in a circle. Protein-coding somatic SNVs and indels are indicated outside each circle as ticks or, for Cancer Gene Census genes, by name. Structural variants (large deletions, translocations, inversions, and duplications) are indicated inside each circle as a line joining the start and end of the variant, with Cancer Gene Census gene mutations indicated outside. Copy number changes are represented in the grey region with higher copy towards the outer edge. Newly appearing SNVs, indels, and structural variants are shown in color, while selected newly appearing CNVs are indicated by colored arrows. CNVs were not determined in the patient #4 day 0 sample (FFPE). Structural variant calls are shown by permission of Aaron Quinlan.



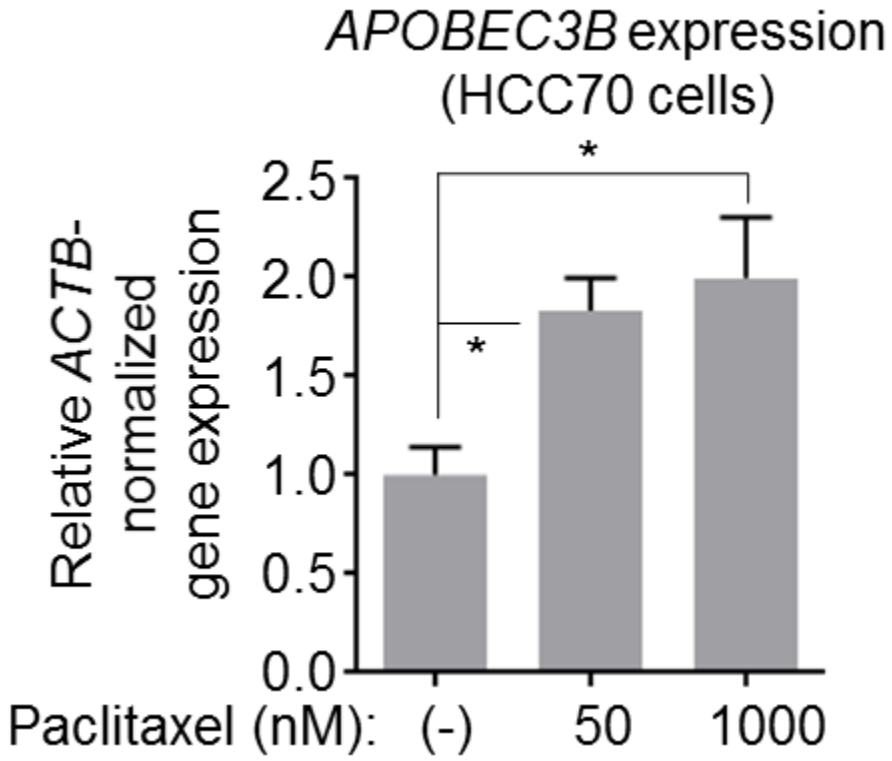
Supplementary Figure 4 | Identification of *SLC25A40-ABCB1* fusion in patient #1 after doxorubicin. (**a**) Structure of *SLC25A40-ABCB1* per WGS (left) and variant allele frequencies over time for the *SLC25A40-ABCB1* fusion (right) in patient #1. (**b**) Schematic of split RNA-Seq reads from post-doxorubicin (day 732) sample. (**c**) Inferred *SLC25A40-ABCB1* fusion transcript (left) and number of RNA-Seq reads (reads spanning plus mate pairs spanning) supporting the *SLC25A40-ABCB1* fusion at each timepoint (right).



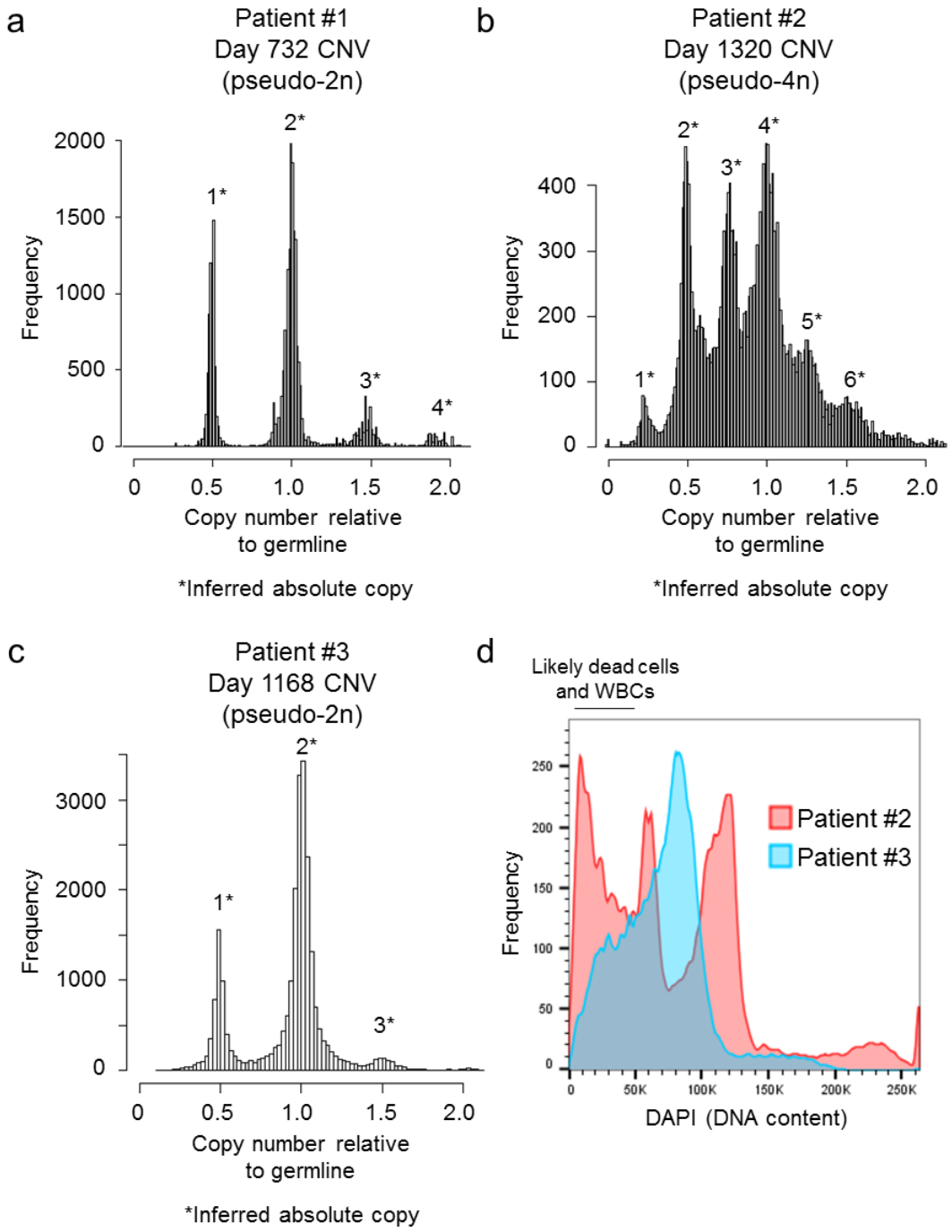
Supplementary Figure 5 | Evolution of *ABCB1* expression in three breast cancer patients. *ABCB1* RPKM (RNA-Seq) for patients #1 (**a**), #2 (**b**), and #3 (**c**) over time with drug treatment indicated in grey. See Fig. 2 legend for drug abbreviations. Made in conjunction with and used with permission of Stephen Piccolo.



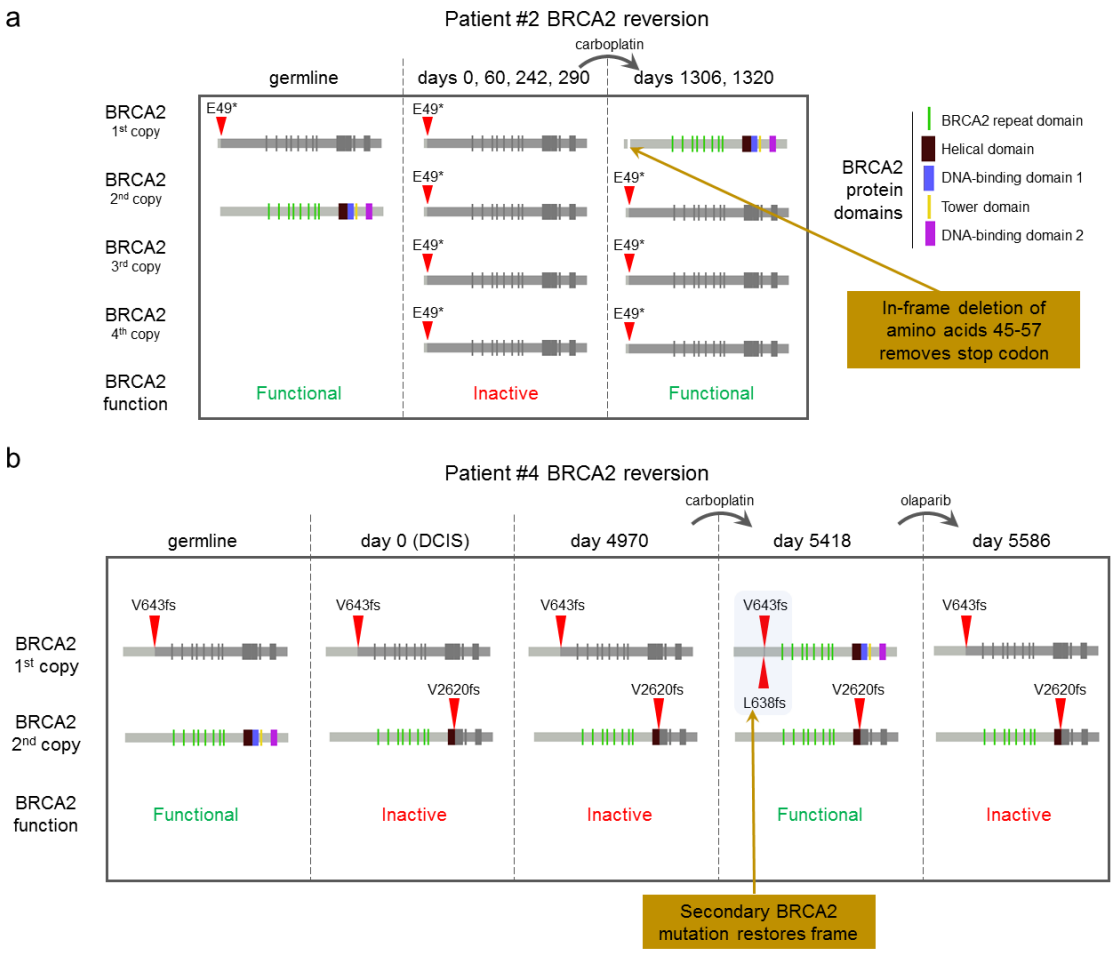
Supplementary Figure 6 | Mutation signature evolution in four breast cancers. Heatmap showing percent of SNVs in each subclone, defined in Fig. 2, with each of 96 possible mutation/trinucleotide context combinations.



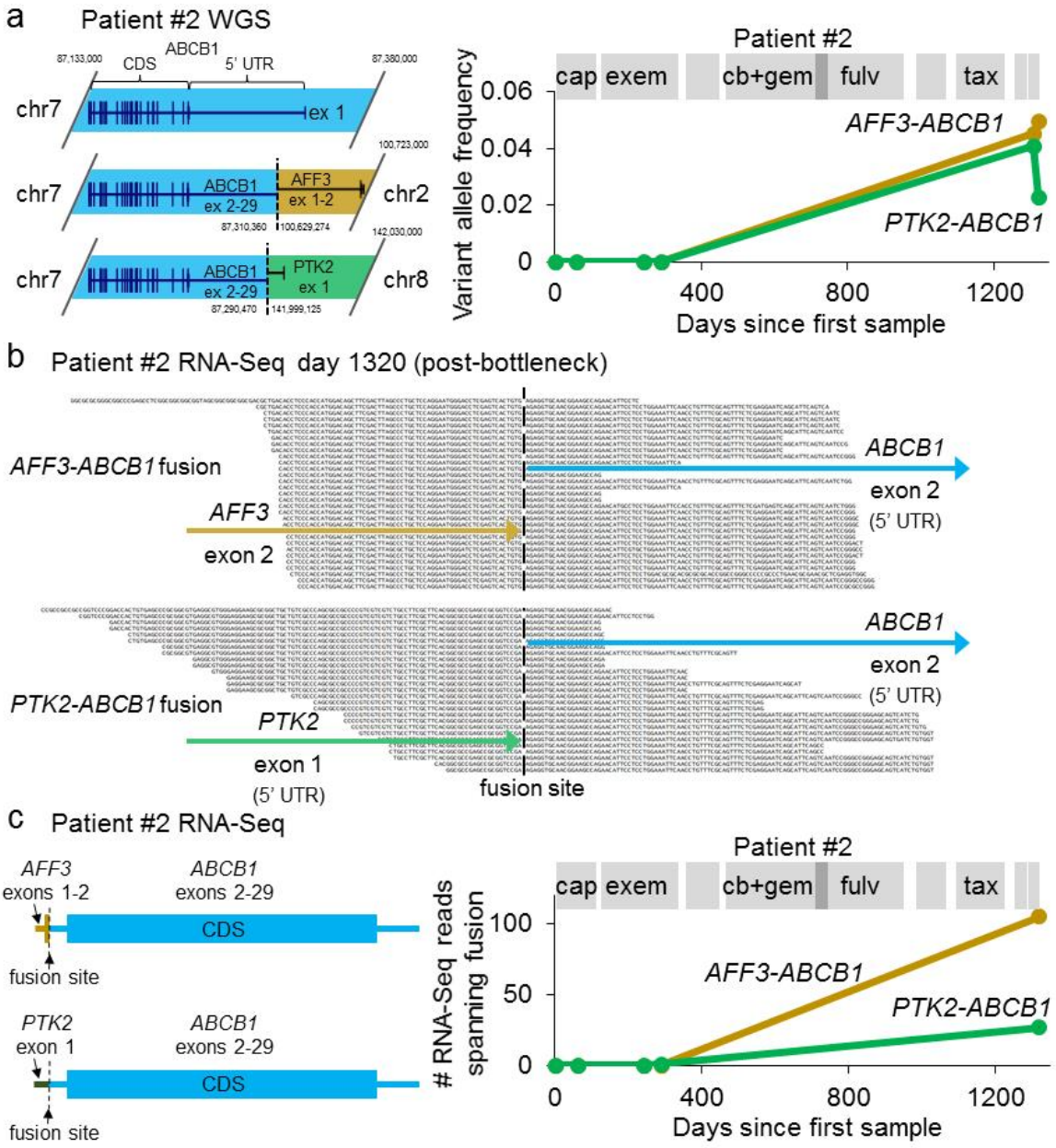
Supplementary Figure 7 | Paclitaxel promotes *APOBEC3B* expression. HCC70 cells were treated 24 hours at indicated doses, followed by qRT-PCR. *APOBEC3B* expression was normalized to *ACTB* expression. * indicates $p < 0.01$ by one-way ANOVA.



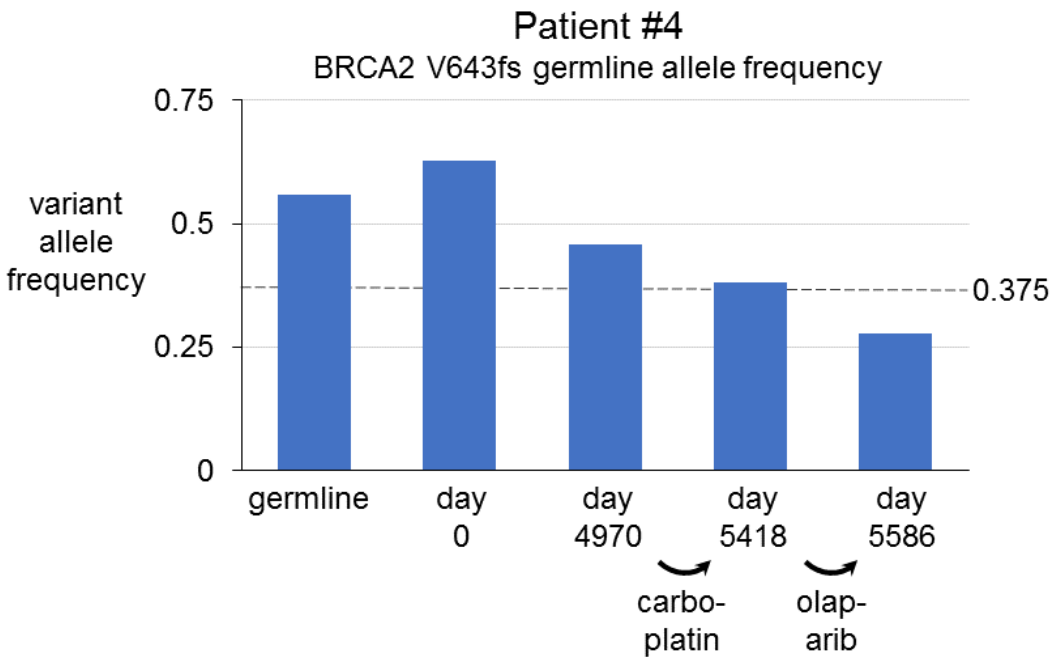
Supplementary Figure 8 | Patient #2 breast cancer is pseudo-tetraploid. Copy number relative to germline for each gene was plotted on a histogram for example samples from patients #1 (**a**), #2 (**b**), and #3 (**c**). Inferred absolute copy is indicated with asterisks by peaks. (**d**) Flow cytometry was performed on example samples from patients #2 (red) and #3 (light blue) using DAPI to quantify DNA content.



Supplementary Figure 9 | *BRCA2* reversions after platinum treatment in two breast cancer patients. Schematics showing mutation status of *BRCA2* in patients #2 (**a**) and #4 (**b**). BRCA2 protein domain structure is from N- (left) to C-terminus (right), with mutations indicated as red arrows. Copy number for *BRCA2* was inferred as described in Methods, and number of copies of each mutation was inferred using mutation allele frequencies.



Supplementary Figure 10 | Identification of *AFF3-ABCB1* and *PTK2-ABCB1* fusions in patient #2 after treatment. (**a**) Structure of *AFF3-ABCB1* and *PTK2-ABCB1* fusions per WGS (left) and their variant allele frequencies over time (right) in patient #2. (**b**) Schematic of split RNA-Seq reads from post-bottleneck (day 1320) sample. Reverse complements of anti-sense reads are shown for clarity. (**c**) Inferred *AFF3-ABCB1* and *PTK2-ABCB1* fusion transcripts (left) and number of RNA-Seq reads (split reads) supporting these fusions at each timepoint (right).



Supplementary Figure 11 | Possible direct BRCA2 reversion in patient #4 at last timepoint. Variant allele frequencies of patient #4 BRCA2 germline mutation (V643fs) at each timepoint.

CHAPTER 3

PHENOTYPIC EVOLUTION OF FOUR BREAST CANCERS AND ALTERNATIVE TREATMENTS

Introduction

Just as understanding the genetic evolution of breast cancer after treatment can suggest treatment options for drug-resistant diseases, elucidating the phenotypic evolution of breast cancers after treatment may suggest alternative treatment options for drug-resistant breast cancer[2]. Further, phenotypic evolution patterns may indicate treatments when no clear drug resistance mutation is identified. Thus, we studied the phenotypic evolution of four breast cancers using both bulk and single-cell RNA-Seq. Based on these data, we identified an effective drug combination in drug-resistant patient cells from one patient.

Materials and Methods

RNA isolation from samples

Cancer samples were obtained and normal cells were depleted as described in the Materials and Methods section of Chapter 2. RNA was isolated using Qiagen RNeasy Micro/Mini Kit and sequenced at HCI or NantOmics using Illumina TruSeq Stranded

mRNA Sample Prep with oligo dT selection or TruSeq Stranded Total RNA Sample Prep Kit with RiboZero Gold library prep and Illumina HiSeq.

RNA-Seq data processing

RNA-Seq data were processed with Rsubread[41,42] v1.16.1 (https://bioconductor.org/packages/release/bioc/html/Rsubread.html) in R using only uniquely mapped reads and the Hamming distance to break ties. The maximum indels allowed per alignment was 5. Gene-level expression values were processed to RPKM (bulk RNA-Seq) or TPM (scRNA-Seq).

Proliferation index

QC-passed reads were aligned to hg19 using MapSplice[43]. Alignment profile was determined by Picard Tools v1.64 (http://broadinstitute.github.io/picard/). Aligned reads were sorted and indexed using SAMtools and translated to transcriptome coordinates then filtered using UBU v1.0 (https://github.com/mozack/ubu). Transcript abundance estimates for each sample were performed using RSEM[44] using the UCSC knownGene transcript and gene definitions. Raw RSEM read counts for all RNA-Seq samples were normalized to the overall upper quartile[45]. Log_2 transformed gene expression estimates were analyzed using the PAM50 algorithm[46] after gene-level adjustments estimated from TCGA samples[47] to produce the proliferation score.

Pathway predictions using ASSIGN

Pathways signatures were first developed by overexpressing IGF-1R, EGFR, K-Ras G12V, c-Raf, Bad, myr-Akt, or GFP control in human mammary epithelial cells (HMECs) grown in MEBM (Lonza) using adenovirus[48]. RNA-Seq data and additional methods describing the generation of the signatures can be found on GEO at GSE73628. RNA-Seq data were adjusted for batch effects using ComBat[49]. Differentially expressed gene lists for each pathway were selected using ASSIGN[50] Bayesian gene selection. Gene lists were then used to estimate pathway activity in each of the breast cancer patient samples analyzed. Molecular Signatures Database gene lists were also used as ASSIGN input in some cases.

Single-cell RNA-Seq

Frozen viable patient pleural effusions were thawed and CD45+ white blood cells, and in some cases fibroblasts (Anti-Fibroblast Microbeads, Miltenyi), were depleted using quadroMACS (Miltenyi). Cells were loaded into a Fluidigm C1 or C1 HT single-cell mRNA-seq chip (for 10-17 μm cell diameter), and single-cell libraries were prepared per manufacturer's instructions. Illumina paired-end sequencing was performed and data were processed to TPM using Rsubread. Cells expressing fewer than 1,700 genes or with fewer than 150,000 mapped reads were excluded. CNV was inferred from scRNA-Seq using the approach described elsewhere[51,52]. Normal human mammary epithelial cells on which scRNA-Seq was performed using Fluidigm C1 were used as the normal (2-copy) state. Mutations were called from scRNA-Seq data by identifying cells with at least 2 reads from subclone-defining mutations. Cells were assigned to subclones if they

possessed CNVs associated with specific subclones, and in the case of patient #1, we verified and refined this based on presence or absence of subclone-defining point mutations in scRNA-Seq. CNVs partially present pre-treatment, and increased post-treatment (i.e. going from absolute copy 2.2 to 3.0), were assumed to belong to the surviving subclone, while those partly present pre-treatment that disappeared were assumed to represent disappearing subclones. Cells were assigned to either survivor or disappearing subclones based on these CNVs. Violin plots of scRNA-Seq ssGSEA scores and TPM were generated with Seurat in R[53].

Heatmaps

RTK heatmaps were generated by adding minimum non-zero RTK RPKM expression value to each RTK and calculating $\log_2$ fold change from day 0. Heatmap was generated in R (ComplexHeatmap).

ssGSEA

ssGSEA was run using GSVA v1.14.1 in R[54] (https://www.bioconductor.org/packages/release/bioc/html/GSVA.html) using RNA-Seq RPKM (bulk) or TPM (single-cell) values and Molecular Signatures Database C2 signatures and custom signatures described. To identify gene sets changing most dramatically in bulk RNA-Seq data, we determined the maximum absolute value change between consecutive timepoint pairs in any patient for each signature. Full EMT signature names described in the text are ANASTASSIOU_CANCER_MESENCHYMAL_TRANSITION_SIGNATURE[55],

JECHLINGER_EPITHELIAL_TO_MESENCHYMAL_TRANSITION_UP[56], and HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION; immune signatures are PHONG_TNF_TARGETS_UP and KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION; proliferation signatures are REACTOME_UNWINDING_OF_DNA, REACTOME_G1_S_SPECIFIC_TRANSCRIPTION[57], SA_REG_CASCADE_OF_CYCLIN_EXPR, KALMA_E2F1_TARGETS[58], and KONG_E2F3_TARGETS[59]. Our custom anti-apoptosis gene set consisted of the 6 anti-apoptotic Bcl-2 family members[60] *BCL2*, *BCL2L1*, *BCL2L2*, *MCL1*, and *BCL2A1*, and the 8 inhibitors of apoptosis (IAP) family members[61] *NAIP*, *BIRC2*, *BIRC3*, *XIAP*, *BIRC5*, *BIRC6*, *BIRC7*, and *BIRC8*. The all-58 RTKs signature was assembled from a review[62]. ASSIGN was used to verify and improve interpretation of evolving signatures identified using ssGSEA.

Western blots

Western blots were performed by scraping cells in cold PBS and lysing pelleted cells in lysis buffer (5 mM EDTA, 150 mM NaCl, 50 mM Tris, pH 8.0, 1% Triton X-100, and 0.1% SDS with protease and phosphatase inhibitor cocktails from Sigma) for 15 minutes on ice. Cleared diluted supernatants were boiled in Laemmli buffer 10 minutes and 20-30 μg of protein per lane was run by SDS-PAGE and transferred to PVDF membrane. Western blot on untreated patient #1 cells was done on never-passaged cells, while elacridar-treated cells were passage 4. For Western blots of patient cells, cells were cultured in Renaissance medium (see “Cell culture and drug assays”).

Immunofluorescence staining

Patient #1 never-cultured pleural effusion or ascites cells were plated at ~3,000 cells/well in 384-well black plates with clear bottom in Renaissance medium (Cellaria) with 5% FBS, 25 ng/mL cholera toxin (Sigma), and 1% antibiotic/antimycotic (Life Technologies); to this was added 20% filtered pleural effusion fluid from another patient. After 3-4 days of culture, cells were washed twice with PBS and fixed in 2% paraformaldehyde (Electron Microscope Sciences) in PBS for 15 minutes. Cells were washed thrice with PBS and permeabilized in 0.1% Triton X-100 in PBS. Permeabilized cells were incubated with primary antibody in 2% bovine serum albumin overnight at 4 °C. Samples were washed twice with PBS containing 0.05% Tween-20, then incubated with Alexa-conjugated secondary antibodies (Thermo Fisher) and DAPI (Invitrogen) for 1 hour of incubation. Samples were washed twice and stored in PBS at 4 °C. Imaging was performed using an automated high-throughput fluorescence microscope imaging system.

Cell culture and drug assays

Drug assays were performed using Renaissance medium, a fibroblast feeder system, or fibroblast feeder system with fibroblasts withdrawn for several passages (as indicated in figure legends). Assays were plated in triplicate or quadruplicate in 96-well or 384-well plates at 2,000 to 10,000 cells/well. Cells were treated 24-72 hours after seeding with drugs (Selleck), and were treated for 72 hours. Viability was determined with Cell Titer-Glo (Promega). Percent cell viability was calculated relative to vehicle mean or median (= 100%).

For trametinib and MK2206 drug assays, we used a fibroblast feeder system modified from that reported previously[63]. We plated 800 irradiated mouse embryonic fibroblasts and 5,600 never-cultured patient-derived cells ("cancer + fibroblast" assay), thawed from frozen viable state, per well in white 384-well plates (quadruplicate). Control wells with 800 fibroblasts per well and no cancer cells were also plated ("fibroblast-only" assay). Medium included 10 μM Y-27632[63]. Two to three days later, drug was given. Three days after drug addition, viability was measured. Cancer + fibroblast signal and associated fibroblast-only signal or cancer + fibroblast minus fibroblast (to obtain cancer-only signal) are shown, as indicated in figure legends. All data were normalized to DMSO treatment for cancer + fibroblasts or cancer + fibroblast minus fibroblast. IC50's for cancer + fibroblast minus fibroblast were generated in GraphPad Prism using regression analysis. Synergy was calculated by Bliss independence. A two-sided t-test was performed to compare actual combination response with expected. Neratinib assay was done using a similar feeder system as above, except that cells were grown without fibroblasts for at least 3-5 passages before assay, so fibroblast response was not subtracted out. Salinomycin, UMI-77 and ABT-263 response assays were performed without fibroblasts in Renaissance Essential Tumor Medium (Cellaria) with 5% FBS, 25 ng/mL cholera toxin (Sigma), and 1% antibiotic/antimycotic (Life Technologies); to this was added 20% filtered pleural effusion fluid from another patient. Patient Westerns were done in Renaissance medium. For TNF-α and TGF-β assays, cells were treated 18 hours with 5 ng/mL TGF-β or 10 ng/mL TNF-α 18, followed by Western blot.

Statistical analysis

scRNA-Seq ssGSEA enrichment scores and *B2M* expression were compared between pre- and post-treatment by student's t-test (two-tailed). Within-sample subclone comparisons were done by one-way ANOVA. A two-proportion test was used to compare the number of scRNA-Seq cells expressing each gene between pre- and post-treatment (two-tailed). P-value plots were generated by plotting $-1*\log_{10}$(p-value) on the y-axis against the ranks for these same values (higher ranks for lower p-values). Error bars for all drug assays represent standard deviation.

Results and Discussion

Tumor phenotype evolution in response to chemotherapy determined by bulk RNA-Seq

To determine how phenotypes evolved through treatment in our patients, we first performed bulk RNA-Seq on longitudinal cancer samples. We then applied ssGSEA[64] to this RNA-Seq data to assign each sample an enrichment score for 3,331 gene sets (signatures) from the Molecular Signatures Database's[65] C2 collection. To identify pathways evolving between longitudinal timepoints, we identified the maximum enrichment score change between consecutive timepoint pairs in any patient for each signature (Supplementary Fig. 12). This revealed that one EMT signature[55] and two proliferation-associated signatures (DNA unwinding and G1/S transcription signatures from Reactome[57]) were among the most dramatically changing signatures between timepoints. Thus, we further investigated EMT and proliferation signature evolution, as well as oncogenic receptor and apoptosis signatures developed in our laboratory (see

Materials and Methods and approach described previously[66]) due to their recognized importance in cancer[67], using ASSIGN[50] to quantify signature strengths in each sample (Fig. 3a). Due to limited material, patient #4's samples were not analyzed by bulk RNA-Seq, but were analyzed via single-cell RNA-Seq as discussed later.

Our analysis revealed increased post-treatment levels of our transcriptional signatures for EGFR, K-Ras G12V (Fig. 3a, under "Signature profiling" in purple), IGF-1R, and Raf (Supplementary Fig. 13), in patients #1 and #2, concomitant with their bottleneck events, suggesting that increased receptor tyrosine kinase (RTK) signaling may have promoted chemotherapy resistance. Numerous individual RTK genes were increased in patients #1 and #2, including *EGFR*, *AXL*, *PDGFRA*, and *AXL*, among others. (See Fig. 3a, right panels under "Individual genes" in purple; see also Supplementary Fig. 14 for expression of all 58 RTKs).

In addition to RTK upregulation, patients #1 and #2 also had increased post-treatment EMT concomitant with their bottleneck events, as indicated by the Anastassiou[55] and Jechlinger[56] EMT signatures from Molecular Signatures Database (Fig. 3a, left panels, in green) and individual EMT genes *VIM* (vimentin), *CDH2* (N-cadherin), *SNAI2* (Snail2), *ZEB1*, and others (Fig. 3a, right panels, in green and Supplementary Fig. 15).

Patient #1 and patient #3 had increased anti-apoptosis after their final treatments. For patient #1, this anti-apoptotic signaling was indicated by an increase in our custom anti-apoptosis signature consisting of anti-apoptotic Bcl-2 and IAP genes (see Materials and Methods) and decrease in our pro-apoptosis Bad signature (Fig. 3a, left, in black). Patient #1 also had increased levels of individual anti-apoptosis genes after carboplatin

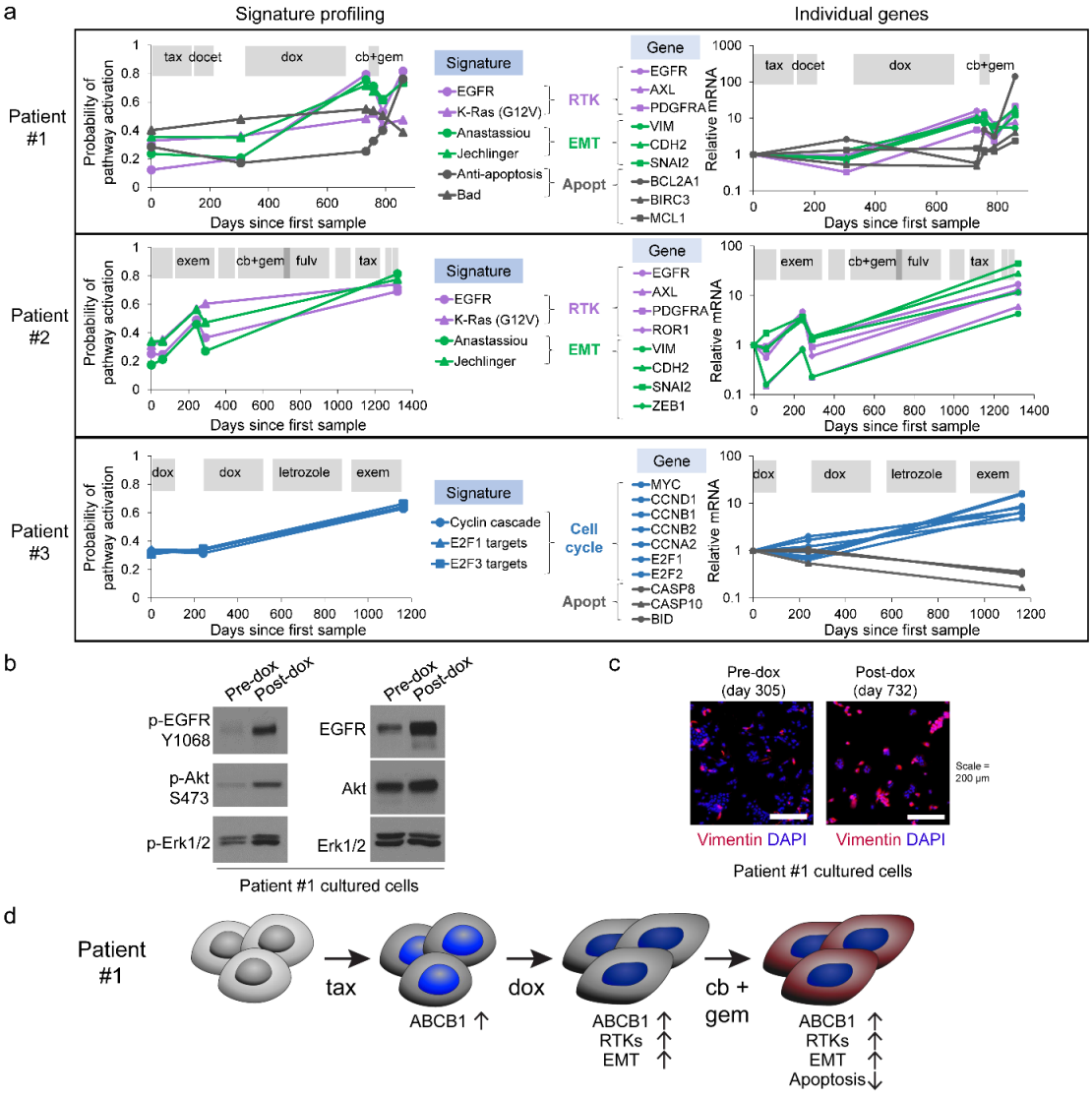


Figure 3 | Phenotypic evolution of three breast cancers including increased post-treatment EMT, RTK, anti-apoptosis, and proliferation properties. ASSIGN signature scores for indicated phenotypes for three breast cancer patients (left side) or day 0-normalized gene RPKM gene expression for indicated genes (right side) based on bulk RNA-Seq data of longitudinal samples. RTK-associated data are in purple, EMT in green, apoptosis in dark grey, and proliferation in blue. Panel (a) was made in conjunction with and used by permission of Evan Johnson, David Jenkins, and Stephen Piccolo.

and gemcitabine, including *BCL2A1* (Bfl-1), *BIRC3* (cIAP2), and *MCL1* (Fig. 3a, right, in black). For patient #3, the anti-apoptotic signaling was indicated through the CNV data (Supplementary Fig. 3c), which showed post-treatment copy loss in regions containing extrinsic apoptosis pathway genes *CASP8, CASP10,* and *BID,* leading to mRNA decrease (Fig. 3a, right, in black).

Patient #3's phenotypic evolution, in contrast to patients #1 and #2, did not involve increased RTK or EMT phenotypes, notwithstanding the acquired *CDH1* (E-cadherin) mutation, consistent with findings that *CDH1* loss-of-function mutations are actually associated with lack of EMT in breast cancer[68]. Instead, this patient's cancer evolved a dramatically increased proliferation phenotype, consistent with our previous analysis indicating that proliferation-associated pathways were among the most dramatically changing between sequential timepoints (Supplementary Fig. 12), after its bottleneck event as shown by increased levels of a cyclin cascade signature and increased E2F1 and E2F3 target signatures[58,59] from the Molecular Signatures Database (Fig. 3a, left, in blue; see Materials and Methods for signature details). Consistent with this, numerous individual cell-cycle-associated genes were increased post-treatment in this patient, including *CCND1* (cyclin D1), *CCNB1* (cyclin B1), *CCNB2* (cyclin B2), *CCNA2* (cyclin A2), *E2F1*, *E2F2*, and *MYC* (Fig. 3a, right, in blue).

We confirmed some of these findings by protein analysis of cultured patient #1 cells. These studies revealed increased post-doxorubicin (day 732) phosphorylation of EGFR, Akt, and Erk1/2 by Western blot (Fig. 3b) and increased post-doxorubicin vimentin by immunofluorescence (Fig. 3c) and Western blot (Supplementary Fig. 16).

Thus, patients #1 and #2 shared similar phenotypic evolution trajectories,

acquiring increased RTK and EMT phenotypes over time, while patient #3 acquired enhanced proliferation and decreased apoptosis. Interestingly, proliferation score analysis indicated that patient #1's proliferation index also evolved significantly, but mostly in response to treatment and not between longitudinal untreated states (Supplementary Fig. 17). This is in contrast to patient #3, whose proliferation changed dramatically between longitudinal untreated states (Fig. 3a). Thus, patient #3's increased proliferation likely represents a true phenotypic change rather than simply being the result of treatment cessation.

The phenotypic changes in patient #1, whose phenotypic evolution was the most complex, is summarized in Fig. 3d, as an example of one patient's step-wise evolution through multiple treatments. After each treatment, patient #1 acquired additional malignant phenotypes likely promoting drug resistance, including increased *ABCB1* expression (Supplementary Fig. 5a) after paclitaxel, increased RTK and EMT phenotypes after doxorubicin, and apoptosis suppression after carboplatin and gemcitabine (Fig. 3d). Interestingly, patient #2 likewise acquired RTK and EMT activation concomitantly, suggesting these two phenotypes may be coordinately up-regulated in some breast cancers after treatment.

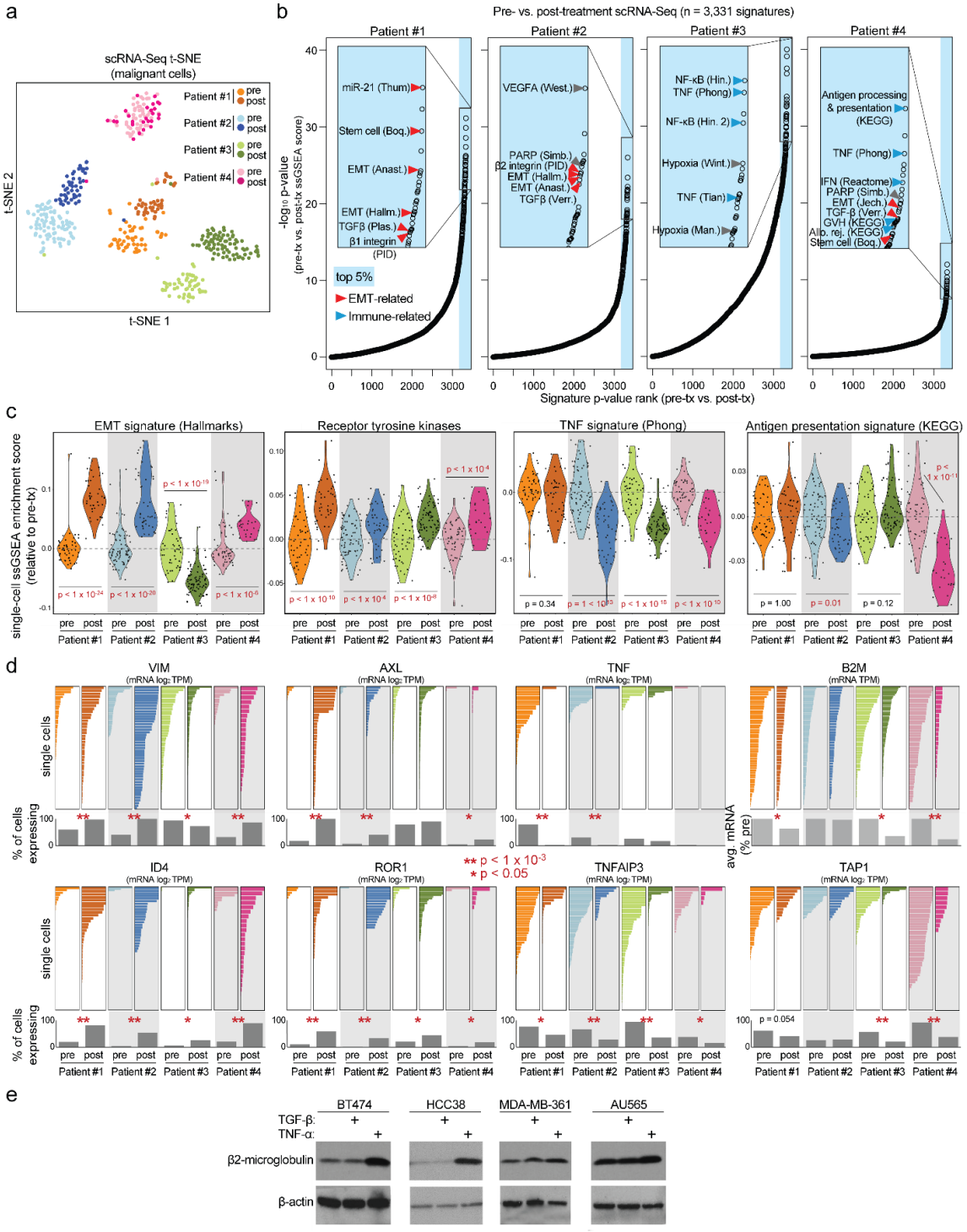
Together, these data highlight the phenotypic plasticity of breast cancer and indicate that EMT and RTK activation may be a recurrent feature of drug-resistant breast cancer. As we later show, these phenotypes can potentially be targeted to better treat drug-resistant breast cancers.

Single-cell RNA-Seq reveals post-treatment immunosuppressive phenotypes

Single-cell RNA-Seq (scRNA-Seq) has several advantages over bulk RNA-Seq, including analysis of intratumoral heterogeneity (genetic and non-genetic) and exclusion of non-malignant cell contaminants that affect bulk measurements[52]. We performed scRNA-Seq on all four patients, at two timepoints each: patient #1 days 305 and 732, patient #2 days 290 and 1320, patient #3 days 249 and 1168, and patient #4 days 4970 and 5586. We identified cells as normal or cancer based on their copy number profiles inferred from scRNA-Seq[52] (Supplementary Figs. 18-21; see also Materials and Methods). This led to a dataset of 428 individual breast cancer cells from 8 specimens.

We used t-SNE analysis[53] to observe global transcriptional differences between malignant cells within and between samples, which showed similarities between samples from the same patient (Fig. 4a). We applied the 3,331 C2 signatures to each single cell using ssGSEA to obtain enrichment scores for each signature, and ranked each signature by its p-value comparing pre- and post-treatment single-cell enrichment scores by student's t-test in each patient (Fig. 4b). This revealed that two broad classes of signatures were most dramatically different between pre- and post-treatment samples: (1) EMT- and stem-cell-associated[69] signatures (red arrows), and (2) immune-associated signatures (blue arrows), including TNF-α and antigen presentation signatures. Single-cell enrichment scores for three highly significant pathways are shown in Fig. 4c, along with a signature for all 58 RTKs[62], which confirmed increased post-treatment EMT and RTK phenotypes, and revealed that TNF-α signaling and antigen presentation decrease after treatment in multiple patients. This suggests that treatment selects for subclones that

Figure 4 | Single-cell RNA-Seq of pre- and post-treatment breast cancers reveals increased immune-avoidance, EMT, and RTK phenotypes. (**a**) t-SNE analysis of 428 individual cells' expression profiles from pre- and post-treatment breast cancer samples from four patients. (**b**) Plots of p-values (t-test-derived) comparing ssGSEA enrichment scores for 3,331 C2 signatures between pre- and post-treatment single cells in each patient. x-axis, p-value ranks (higher ranks are more significant); y-axis, $-\log_{10}$(p-value). Red arrows, EMT- and stem cell-related signatures; blue arrows, immune-related signatures. (**c**) ssGSEA enrichment score (relative to pre-treatment average) violin plots for single cells in each sample for indicated signatures. Each point represents a single cell. p-values are by student's t-test. (**d**) Expression of indicated genes by scRNA-Seq (x-axis) in individual cells (y-axis) with percent of cells expressing each gene or average expression indicated below in grey bars. p-values are by two-sample proportion test. x-axis scales all begin at zero (left) and are the same within patients for each gene. (**e**) Western blot of indicated cells treated for 18 hours with TGF-β (5 ng/mL) or TNF-α (10 ng/mL). Panel (e) was generated in conjunction with and used by permission of Gajendra Shrestha.


a
scRNA-Seq t-SNE
(malignant cells)
Patient #1
Patient #2
Patient #3
Patient #4
pre
post
t-SNE 2
t-SNE 1
b
Pre- vs. post-treatment scRNA-Seq (n = 3,331 signatures)
Patient #1
Patient #2
Patient #3
Patient #4
-log10 p-value
(pre-tx vs. post-tx ssGSEA score)
miR-21 (Thum)
Stem cell (Boq.)
EMT (Anast.)
EMT (Hallm.)
TGFβ (Plas.)
β1 integrin
(PID)
top 5%
EMT-related
Immune-related
VEGFA (West.)
PARP (Simb.)
β2 integrin (PID)
EMT (Hallm.)
EMT (Anast.)
TGFβ (Verr.)
NF-κB (Hin.)
TNF (Phong)
NF-κB (Hin. 2)
Hypoxia (Wint.)
TNF (Tian)
Hypoxia (Man.)
Antigen processing
& presentation
(KEGG)
TNF (Phong)
IFN (Reactome)
PARP (Simb.)
EMT (Jech.)
TGF-β (Verr.)
GVH (KEGG)
Allo. rej. (KEGG)
Stem cell (Boq.)
Signature p-value rank (pre-tx vs. post-tx)
c
EMT signature (Hallmarks)
Receptor tyrosine kinases
TNF signature (Phong)
Antigen presentation signature (KEGG)
single-cell ssGSEA enrichment score
(relative to pre-tx)
p < 1 x 10-24
p < 1 x 10-20
p < 1 x 10-19
p < 1 x 10-6
p < 1 x 10-10
p < 1 x 10-4
p < 1 x 10-8
p < 1 x 10-4
p = 0.34
p < 1 x 10-13
p < 1 x 10-15
p < 1 x 10-10
p = 1.00
p = 0.01
p = 0.12
p < 1 x 10-11
pre
post
Patient #1
Patient #2
Patient #3
Patient #4
d
VIM
AXL
TNF
B2M
ID4
ROR1
TNFAIP3
TAP1
(mRNA log2 TPM)
single cells
% of cells expressing
avg. mRNA (% pre)
** p < 1 x 10-3
* p < 0.05
p = 0.054
e
BT474
HCC38
MDA-MB-361
AU565
TGF-β:
TNF-α:
β2-microglobulin
β-actin

avoid immune responses, as loss of both TNF-α signaling[70,71] and antigen presentation[72] are likely to suppress basal or treatment-induced immune responses. To identify individual genes driving these phenotypes, we statistically analyzed the proportion of single cells expressing each gene between pre- and post-treatment samples in each patient (Supplementary Fig. 22). This revealed, in multiple patients, significant differences between pre- and post-treatment proportions of cells expressing the EMT- and stem cell-associated genes *VIM* (vimentin[56]) and *ID4*[73,74]; the RTK genes *AXL* and *ROR1*[75]; the TNF-α pathway genes *TNF* (TNF-α) and *TNFAIP3*[76]; and the antigen processing gene *TAP1*[77] (Fig. 4d). *EGFR* also increased in post-treatment cells by scRNA-Seq (Supplementary Fig. 23), consistent with bulk RNA-Seq (Fig. 3a, b). In addition, the essential antigen presentation gene *B2M*, encoding the β2-microglobulin subunit required for HLA class I cell-surface expression[78], was present in all cells but had statistically decreased expression by student's t-test in 3 of 4 patients (Fig. 4d, top-right).

Given concurrent EMT increase and loss of TNF-α and antigen presentation after treatment in some patients, we hypothesized that EMT may suppress antigen presentation and that TNF-α may promote antigen presentation. To test this hypothesis, we treated four breast cancer cell lines with the EMT promoter TGF-β[79] or with TNF-α and measured β2-microglobulin by Western blot. While TGF-β had little effect, TNF-α substantially increased β2-microglobulin levels in two cell lines (Fig. 4e). This indicates that TNF-α, in addition to promoting T cell activation, may also increase the susceptibility of cancer cells to cytotoxic T cells by increasing antigen presentation. This also suggests that decreased post-treatment *B2M* levels (Fig. 4d) may be due to concurrent decreased TNF-α signaling (Fig. 4c, d). This immunophenotypic evolution

may have been less apparent in bulk RNA-Seq due to dramatically different (often increased) expression of immune-related genes in normal cells, such as fibroblasts, present in the cancer milieu (Supplementary Fig. 24).

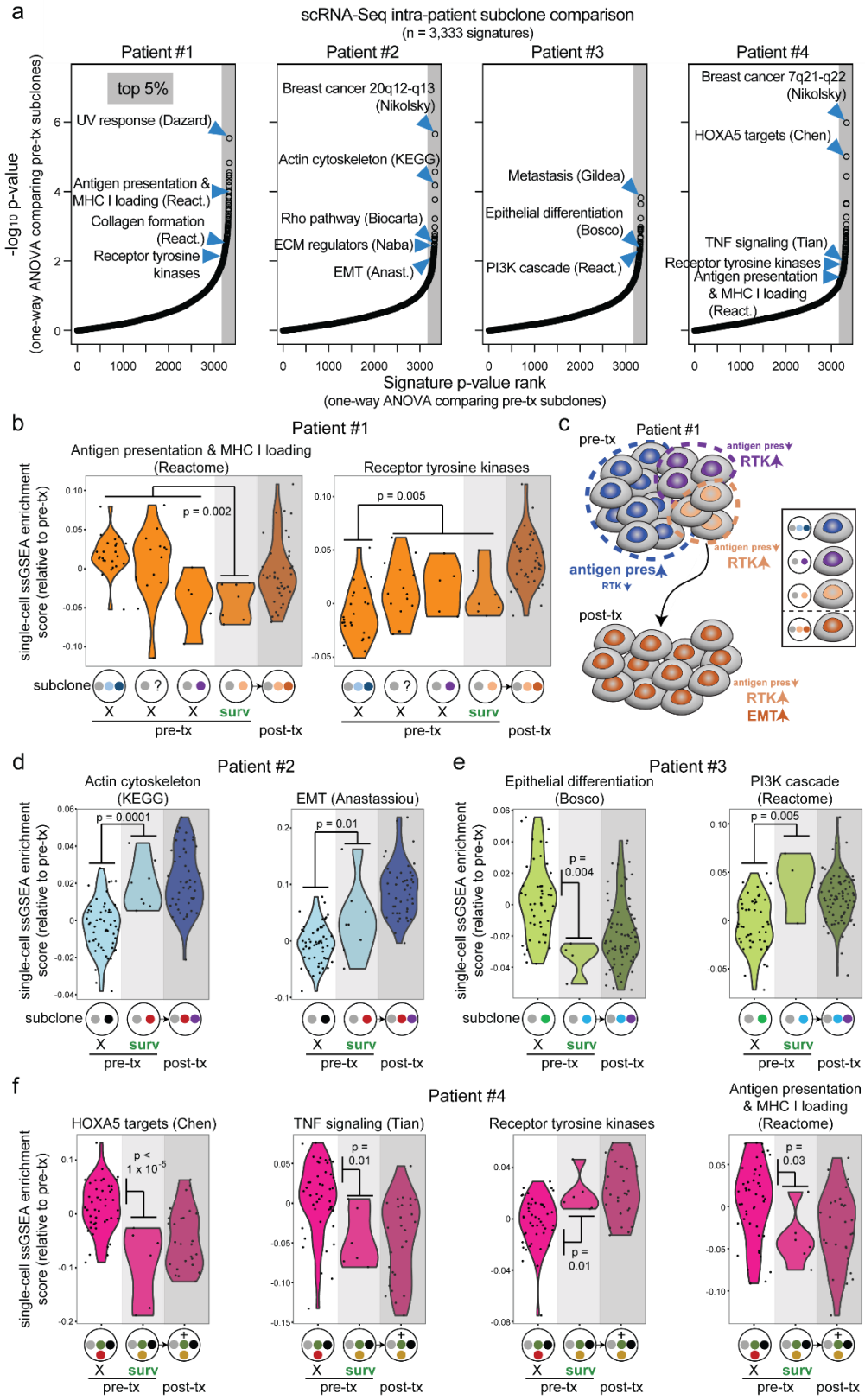
Single-cell RNA-Seq reveals pre-treatment EMT, RTK, and immune-avoidance phenotypes in survivor subclones

Our analysis shows that cancers acquire additional malignant phenotypes in response to therapy. This likely occurs through (1) genetic selection for subclones with these features and/or (2) drug-induced changes. To test whether acquired phenotypes were present before treatment in survivor subclones (scenario #1), we assigned pre-treatment single cells to subclones identified using DNA sequencing. To assign each cell to a subclone, we used a combination of CNV inferred from scRNA-Seq[52] data and analysis of subclone-defining point mutations with scRNA-Seq coverage (Supplementary Figs. 18-21; "*" indicates relevant CNVs). CNVs that were partially present before treatment, and subsequently increased post-treatment, were assumed to represent a pre-treatment *survivor* subclone, while those that were partially present pre-treatment but later disappeared were assumed to represent a *disappearing* subclone. Cells were assigned to either survivor or disappearing subclones based on these CNVs; subclone assignment was further refined by identification of rare cells with coverage of subclone-defining mutations (Supplementary Fig. 18).

To identify signatures significant different between pre-treatment subclones, we compared ssGSEA enrichment scores for 3,331 C2 signatures, plus the all-58 RTKs signature[75] and our anti-apoptosis signature, between pre-treatment subclones by one-way

ANOVA for each patient (Fig. 5a). This revealed differences in antigen presentation and RTK signatures between subclones in patient #1 (Fig. 5a and 5b). The patient #1 pre-treatment survivor subclone that gave rise to the post-treatment bottleneck subclone had decreased antigen presentation and increased RTK expression compared to other subclones, suggesting that these phenotypes were genetic and pre-existed before treatment. Thus, in patient #1, several pre-treatment subclones varied in their antigen presentation and RTK levels (Fig. 5c). A subclone with decreased antigen presentation and increased RTK expression (light-orange) may have had a selective advantage and thus gave rise to the post-treatment bottleneck subclone (dark-orange), which had additionally evolved increased EMT (Fig. 5c). Further, the patient #2 survivor subclone also had a partially increased EMT- and actin cytoskeleton-associated gene expression, compared to the disappearing subclone, while post-treatment EMT levels increased further (Fig. 5d), although assignment of single cells to subclones in this patient was less clear due to dependence on a single CNV (on chromosome 20). The patient #3 survivor subclone possessed a diminished epithelial differentitation phenotype and an enhanced PI3K signature that was apparently carried into its descendant subclone after treatment (Fig. 5e). Finally, the patient #4 survivor subclone possessed decreased expression of transcriptional targets of the *HOXA5* tumor suppressor[80], decreased TNF signaling, decreased antigen presentation, and increased RTK expression, which were likewise present in its descendant post-treatment subclones (Fig. 5f). These data indicate that the post-treatment RTK, EMT, and immune-avoidance phenotypes pre-existed before treatment in a genetically defined subclone in some cases, and may have been selected for on a genetic basis.

Figure 5 | Pre-existence of post-treatment phenotypes in pre-treatment survivor subclones. (**a**) Plots of p-values (one-way ANOVA) comparing scRNA-Seq ssGSEA enrichment scores for 3,331 C2 signatures, plus the receptor tyrosine kinases (all-58) signature and our anti-apoptosis signature, between subclones in each patient's pre-treatment sample. x-axis, p-value ranks (higher ranks are more significant); y-axis, $-\log_{10}$(p-value). (**b**) ssGSEA enrichment scores (or individual gene expression levels) for single cells in each pre-treatment subclone, with post-treatment cells shown for comparison. Each dot represents a single cell. p-values are by student's t-test. Subclones correspond to those shown in Fig. 2a. "X" indicates disappearing subclone while "surv" indicates the subclone giving rise to the post-treatment sample. (**c**) Schematic showing subclonal phenotypic heterogeneity in pre-treatment patient #1 cells. (**d**) As in (b) but for patient #2. Subclones correspond to those in Fig. 2b.


a
scRNA-Seq intra-patient subclone comparison
(n = 3,333 signatures)
Patient #1
Patient #2
Patient #3
Patient #4
top 5%
UV response (Dazard)
Antigen presentation & MHC I loading (React.)
Collagen formation (React.)
Receptor tyrosine kinases
Breast cancer 20q12-q13 (Nikolsky)
Actin cytoskeleton (KEGG)
Rho pathway (Biocarta)
ECM regulators (Naba)
EMT (Anast.)
Metastasis (Gildea)
Epithelial differentiation (Bosco)
PI3K cascade (React.)
Breast cancer 7q21-q22 (Nikolsky)
HOXA5 targets (Chen)
TNF signaling (Tian)
Receptor tyrosine kinases
Antigen presentation & MHC I loading (React.)
-log10 p-value
(one-way ANOVA comparing pre-tx subclones)
Signature p-value rank
(one-way ANOVA comparing pre-tx subclones)
b
Patient #1
Antigen presentation & MHC I loading (Reactome)
Receptor tyrosine kinases
p = 0.002
p = 0.005
single-cell ssGSEA enrichment score (relative to pre-tx)
subclone
surv
pre-tx
post-tx
c
Patient #1
pre-tx
antigen pres
RTK
EMT
post-tx
d
Patient #2
Actin cytoskeleton (KEGG)
EMT (Anastassiou)
p = 0.0001
p = 0.01
e
Patient #3
Epithelial differentiation (Bosco)
PI3K cascade (Reactome)
p = 0.004
p = 0.005
f
Patient #4
HOXA5 targets (Chen)
TNF signaling (Tian)
Receptor tyrosine kinases
Antigen presentation & MHC I loading (Reactome)
p < 1 x 10-5
p = 0.01
p = 0.01
p = 0.03

Targeting evolving phenotypes in tumors using adaptive treatment strategies

It is important to identify adaptive therapeutic regimens that effectively target the dynamic, heterogeneous nature of breast cancer. Even where it is not possible to target a specific mutation, we can match treatments to altered cellular phenotypes. Because RTK activation and the EMT phenotype increased after treatment-induced bottleneck events in three patients (Figs. 3a, 4c), we hypothesized that these post-treatment cells may have acquired sensitivity to drugs targeting these phenotypes. To test this hypothesis, we compared the response of patient #1 pre- and post-doxorubicin cancer cells (day 305 and 732) to inhibitors of MEK and AKT (trametinib[81] and MK2206[82]), two important RTK effectors[62], and the HER/ErbB family (neratinib[83]), given this patient's early classification as HER2+ per clinicians. We also tested the EMT and cancer stem cell inhibitor salinomycin[69,84], due to enhanced post-doxorubicin EMT. Patient-derived breast cancer cells were grown using a modified fibroblast feeder approach[63] and other approaches (see Materials and Methods).

Post-doxorubicin patient #1 cells were moderately more sensitive to MEK and AKT inhibitors than pre-doxorubicin cells (Fig. 6a). We also tested these drugs in combination due to redundancy between PI3K and MAPK signaling[85–87]. Remarkably, post-doxorubicin cells were dramatically more sensitive to the equimolar combination of trametinib and MK2206 (Fig. 6b, top), with a combined IC50 of 4.37 μM in the post-doxorubicin cells and an unmet IC50 in the pre-doxorubicin cells due to lack of response at doses tested (see also Supplementary Fig. 25a). The combination was synergistic at 3.125 μM (Fig. 6b, bottom) and 6.25 μM (Supplementary Fig. 25b), using Bliss independence, in the post-doxorubicin but not the pre-doxorubicin cells. Neratinib and

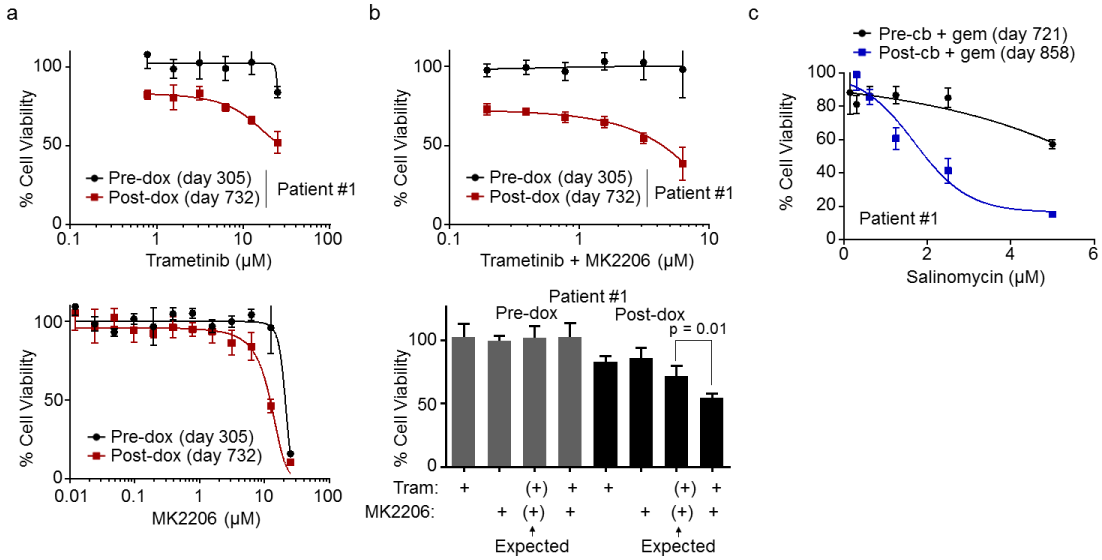


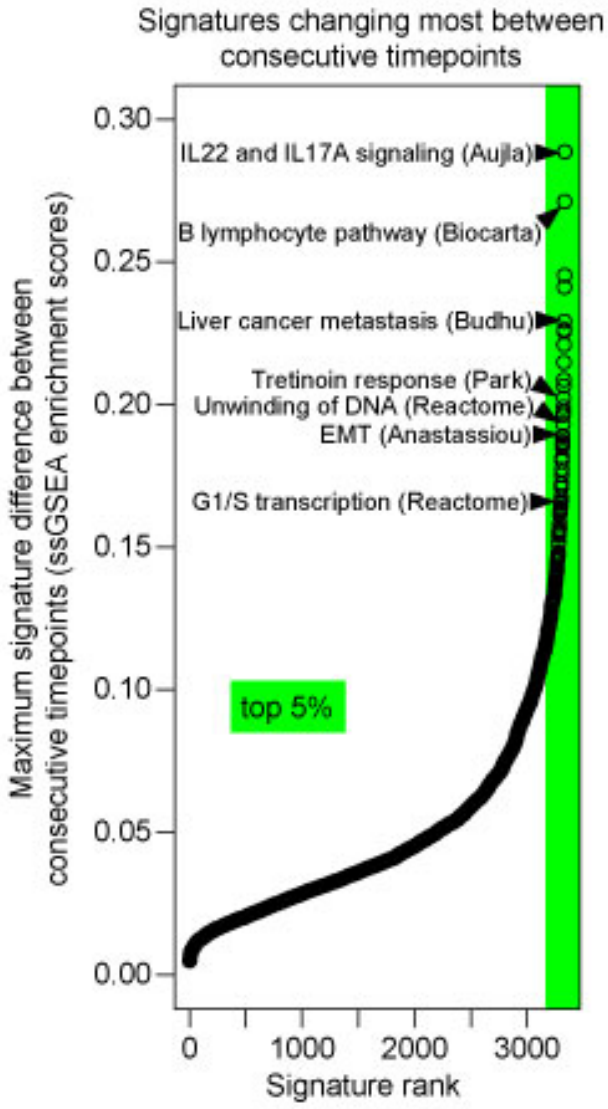
Figure 6 | Acquired sensitivity to drugs targeting post-chemotherapy phenotypes. (**a**) Drug response assay comparing pre- and post-doxorubicin patient #1 cancer cells' sensitivity to drugs after 3-day treatment (fibroblast feeder system; CellTiter-Glo was used). (**b**) Indicated cells were treated with equimolar doses of trametinib and MK2206 for 3 days, followed by CellTiter-Glo (top); synergy analysis of 3.125 μM doses of these drugs (bottom). Expected is by Bliss independence; p-value is by student's t-test. Fibroblast feeder system was used and fibroblast signal was subtracted out in (a) and (b). (**c**) Drug response assay comparing patient #1 pre- and post-carboplatin and gemcitabine cancer cells' sensitivity to indicated drugs. Renaissance medium was used.

salinomycin showed equivalent responses between pre- and post-doxorubicin cells (Supplementary Fig. 26a). These data suggest that enhanced RTK signaling post-doxorubicin (Fig. 3a, b) induced increased dependency on downstream PI3K and MAPK signaling. We also found that patient tumor cells taken post-carboplatin and gemcitabine, while not more sensitive to Bcl-2 family inhibitors (Supplementary Fig. 26b) notwithstanding increased Bcl-2 family gene expression (Fig. 3a), were more sensitive to salinomycin than pre-treatment cells (Fig. 6c). This response correlates with increased EMT, including decreased E-cadherin (*CDH1*), after carboplatin and gemcitabine (Fig. 3a, Supplementary Fig. 15). Interestingly, there was no differential salinomycin response between pre- and post-doxorubicin timepoints, perhaps due to the cells' dependence on growth receptor signaling after doxorubicin (Supplementary Fig. 26a).

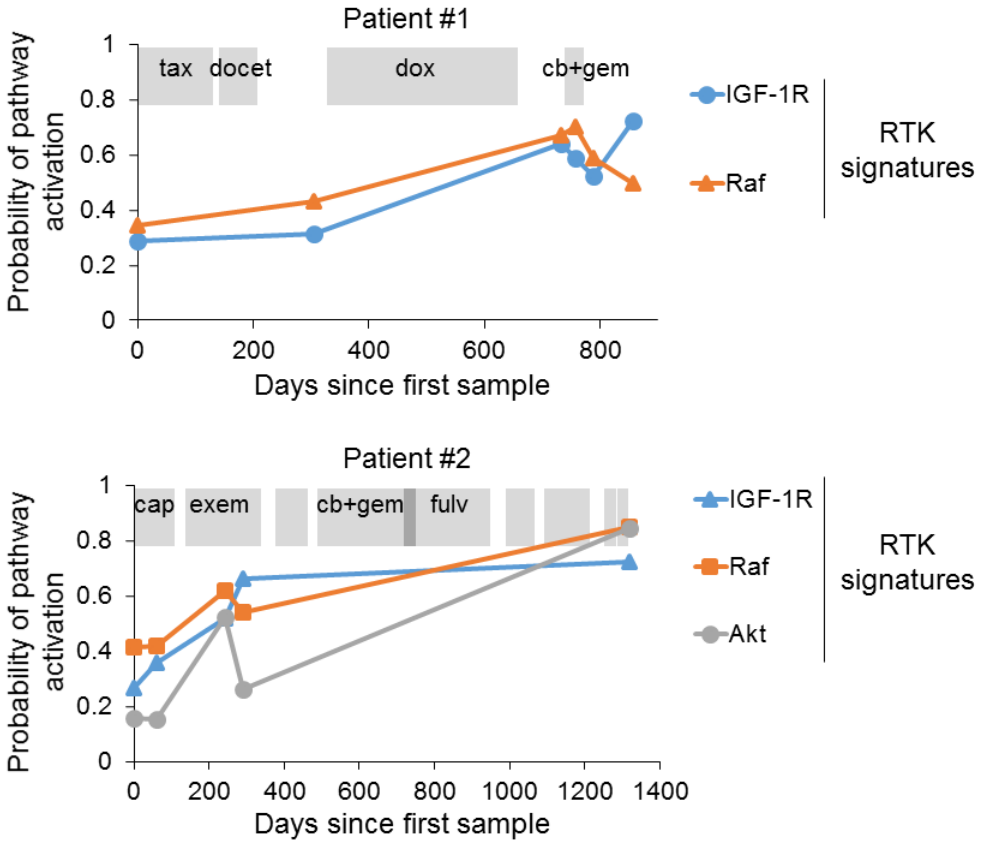
These findings indicate that the development of disease refractory to one treatment may be associated with enhanced sensitivity to an alternative treatment targeting post-treatment emerging phenotypes. This approach can potentially be employed in drug-resistant cancers to improve outcomes.

Supplemental Material

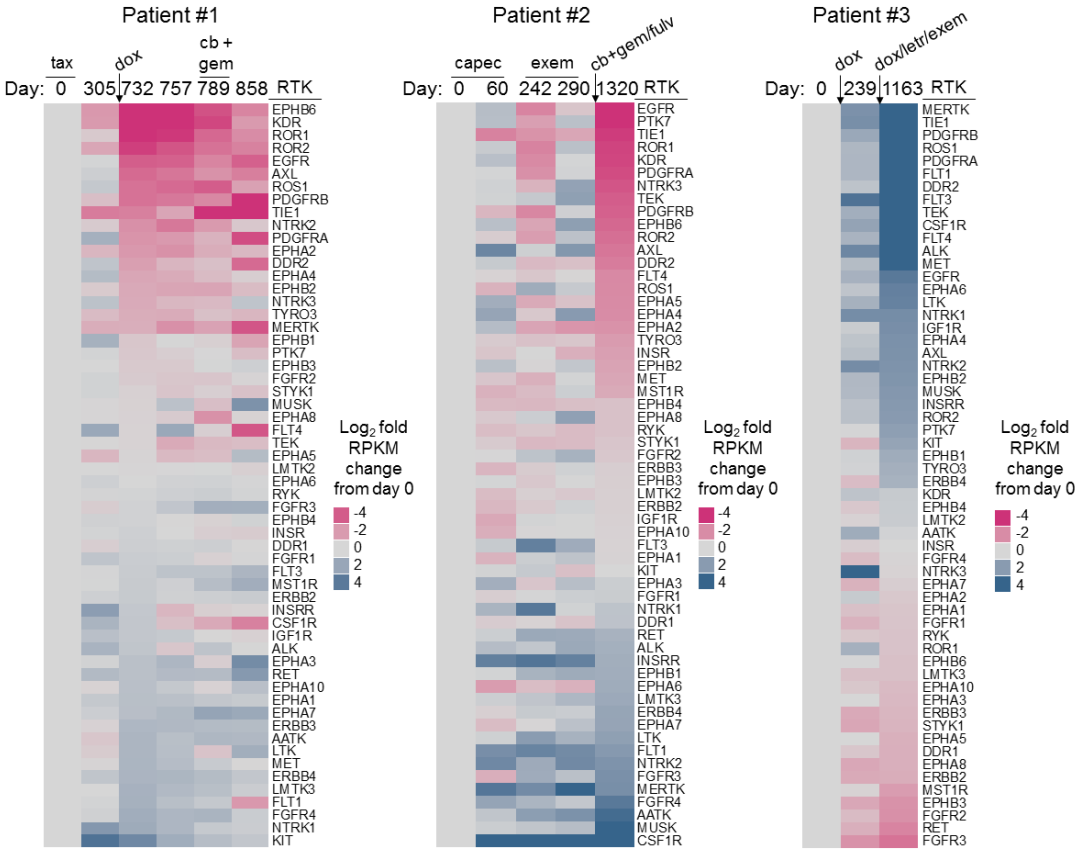
Supplementary figures referred to in this chapter can be found on the ensuing pages.



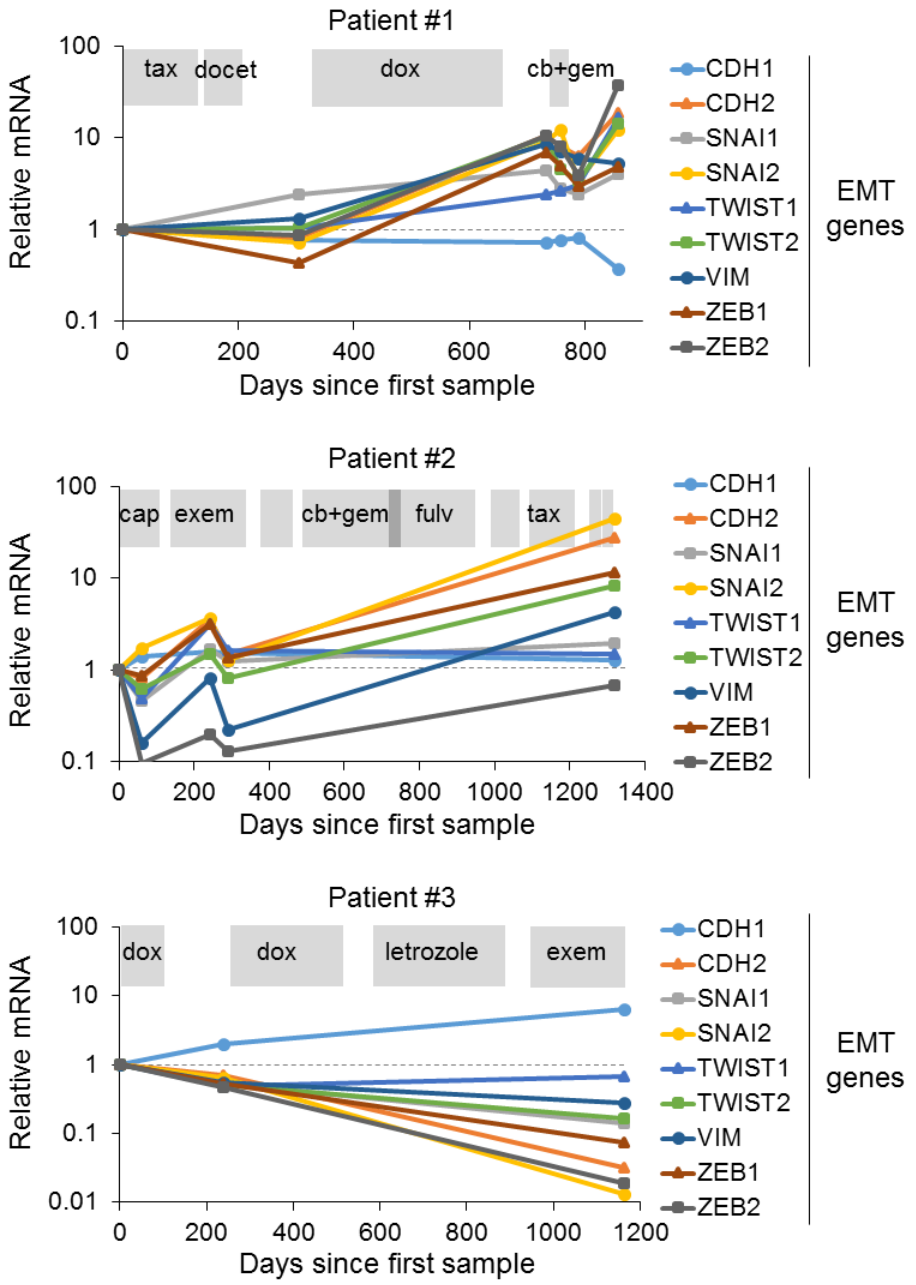
Supplementary Figure 12 | Identification of dramatically evolving signatures in bulk RNA-Seq data from three breast cancers. ssGSEA enrichment scores were determined for bulk RNA-Seq samples for patients #1-#3 for 3,331 C2 signatures, and the maximum absolute value difference between consecutive timepoints from any patient were plotted (y-axis) against their ranks (x-axis; higher rank indicates greater difference).



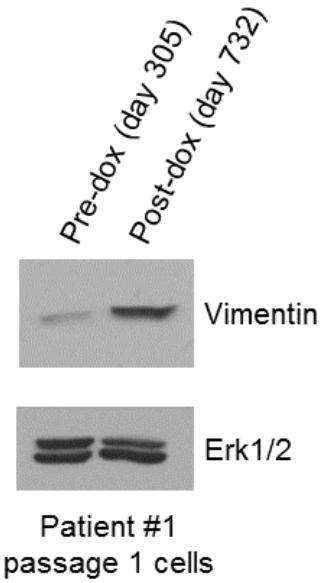
Supplementary Figure 13 | Additional RTK signatures increasing in two breast cancer patients after treatment. ASSIGN signature scores for indicated pathways for two breast cancer patients. Expression signatures were generated by experimentally overexpressing indicated genes in human mammary epithelial cells, followed by RNA-Seq to identify genes indicative of pathway activation. Made in conjunction with and used by permission of Stephen Piccolo, Evan Johnson, and David Jenkins.



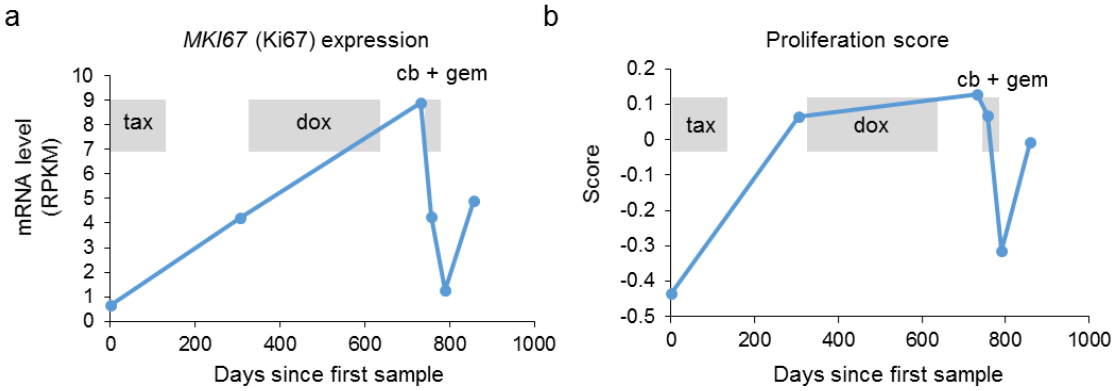
Supplementary Figure 14 | Evolution of RTK gene expression in three breast cancer patients. Heatmaps showing $\log_2$ fold change expression from day 0 (RPKM expression values) for all 58 receptor tyrosine kinases over time in breast cancer patients #1 (left), #2 (middle), and right (#3). RTKs are sorted by fold-change expression from day 0 in the post-bottleneck timepoint for each patient.



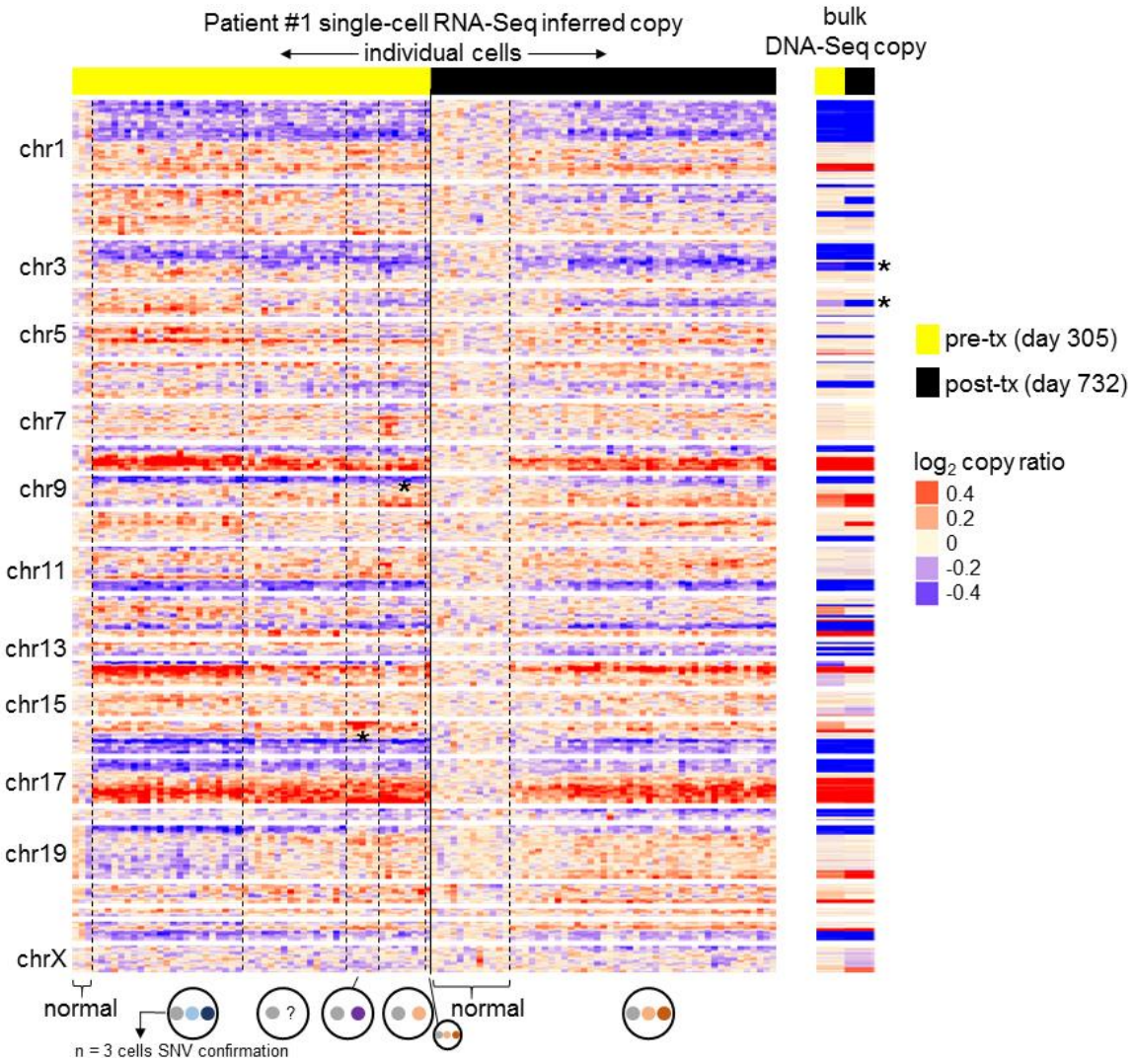
Supplementary Figure 15 | Evolution of EMT-associated gene expression in three breast cancer patients. Day 0-normalized expression (RNA-Seq RPKM values) for EMT-associated transcription factors and other genes for patients #1 (top), #2 (middle), and #3 (bottom) over time with drug treatment indicated in grey. See Fig. 2 legend for drug abbreviation definitions.



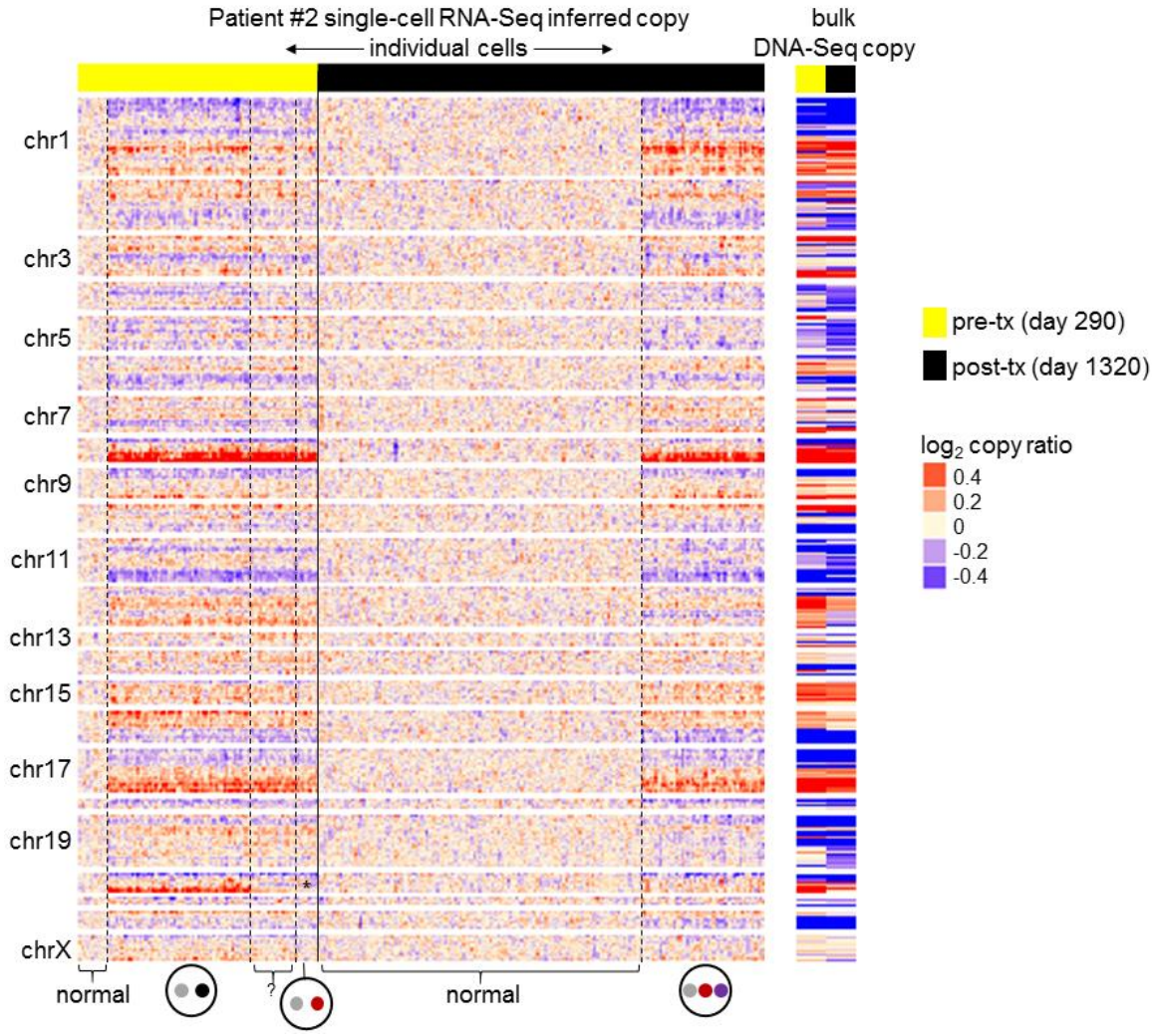
Supplementary Figure 16 | Vimentin is increased in patient #1 cells per Western blot. Western blot of cultured patient #1 pre- (day 305) and post-doxorubicin (day 732) cells using same lysate as in Fig. 3b and same Erk1/2 loading control. Renaissance medium was used (see Methods).



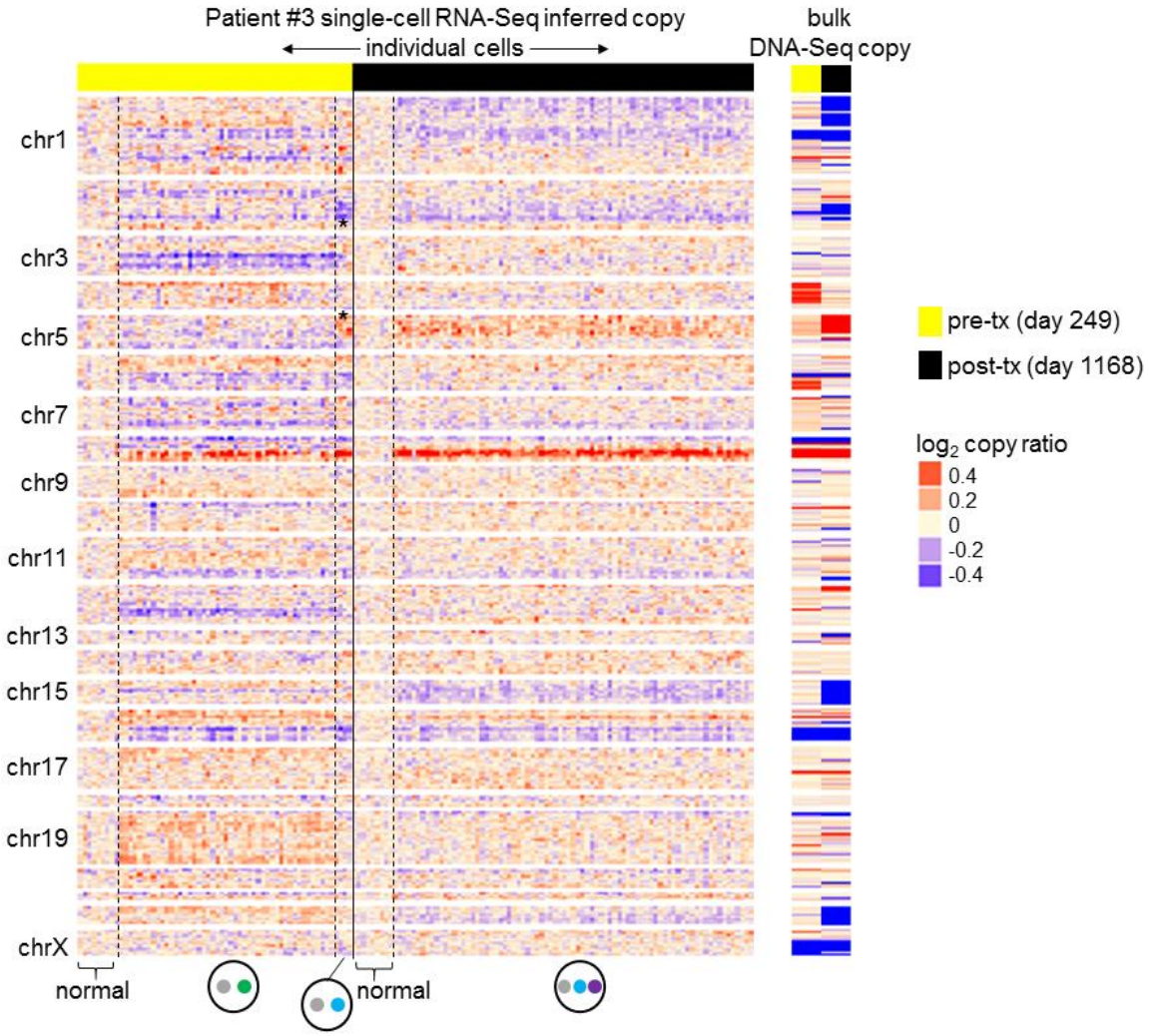
Supplementary Figure 17 | Changes in proliferation over time in patient #1. (**a**) Expression of the gene encoding Ki67 per RNA-Seq (RPKM) over time. (**b**) Proliferation score over time in patient #1. (b) was made in conjunction with and used with permission of Joel Parker and Sara Selitsky.



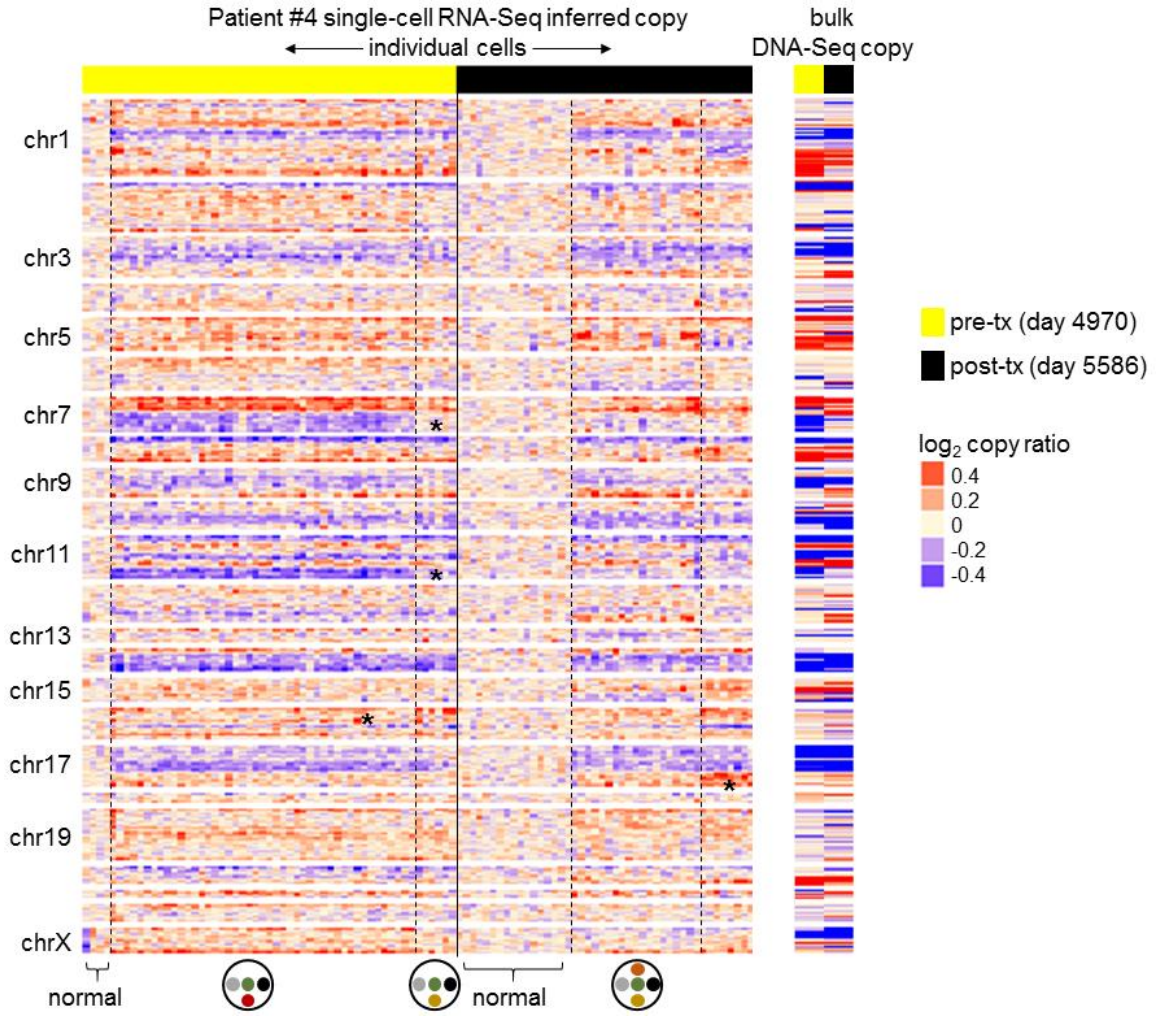
Supplementary Figure 18 | Copy number and subclones inferred from single-cell RNA-Seq of patient #1. scRNA-Seq was performed on pre- and post-treatment cells from patient #1 and copy number was inferred and corroborated with bulk DNA-Seq-based copy for each gene. Subclones were identified using CNVs and subclone-defining point mutations as described in text and Methods. * indicates CNVs aiding with subclone identification. ? indicates cells with unclear subclone identity. Subclones assigned (bottom) are defined as in Fig. 2a.



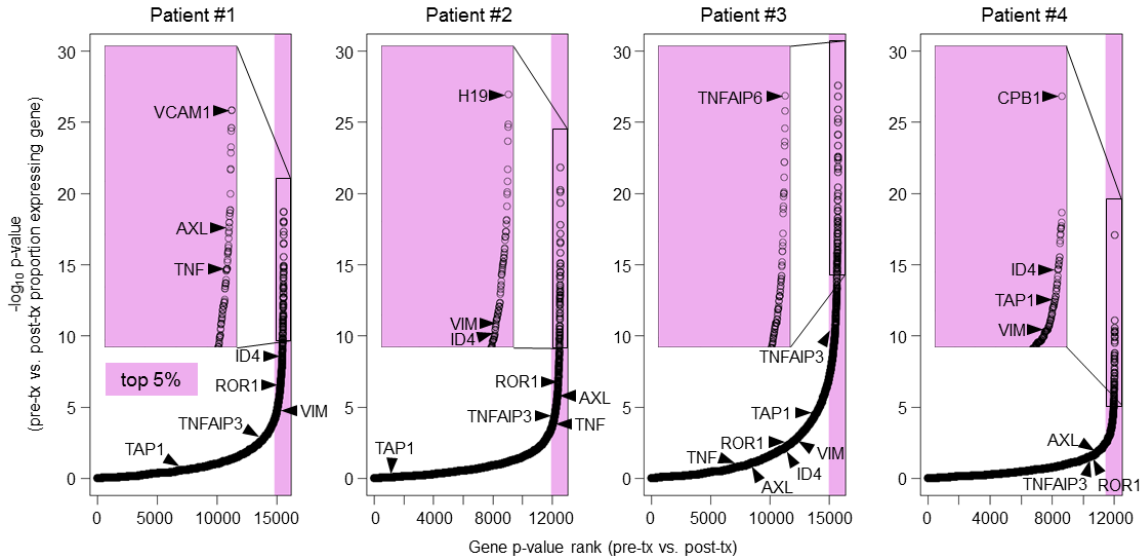
Supplementary Figure 19 | Copy number and subclones inferred from single-cell RNA-Seq of patient #2. scRNA-Seq was performed on pre- and post-treatment cells from patient #2 and copy number was inferred and corroborated with bulk DNA-Seq-based copy for each gene. Subclones were identified using CNV (chromosome 20q loss) as described in text and Methods. * indicates CNV aiding with subclone identification. ? indicates cells with unclear subclone identity (equivocal 20q status). Subclones assigned (bottom) are defined as in Fig. 2b.



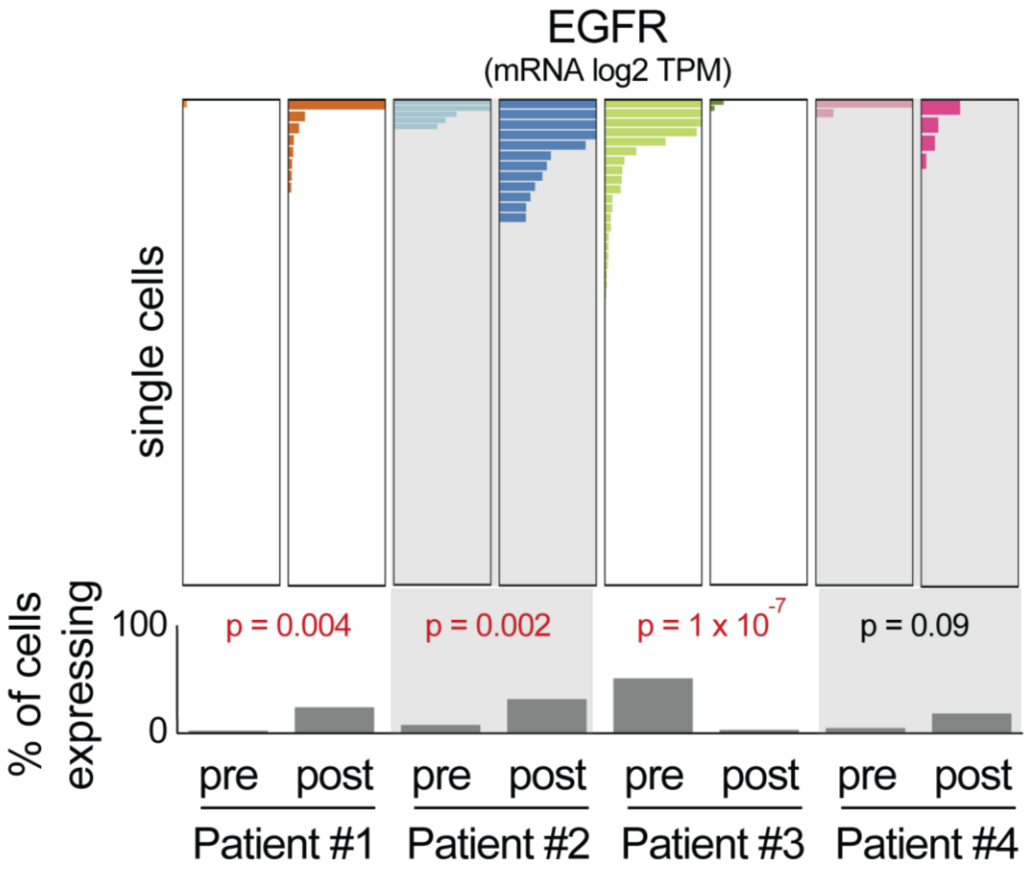
Supplementary Figure 20 | Copy number and subclones inferred from single-cell RNA-Seq of patient #3. scRNA-Seq was performed on pre- and post-treatment cells from patient #3 and copy number was inferred and corroborated with bulk DNA-Seq-based copy for each gene. Subclones were identified using CNVs as described in text and Methods. * indicates CNVs aiding with subclone identification. Subclones assigned (bottom) are defined as in Fig. 2c.



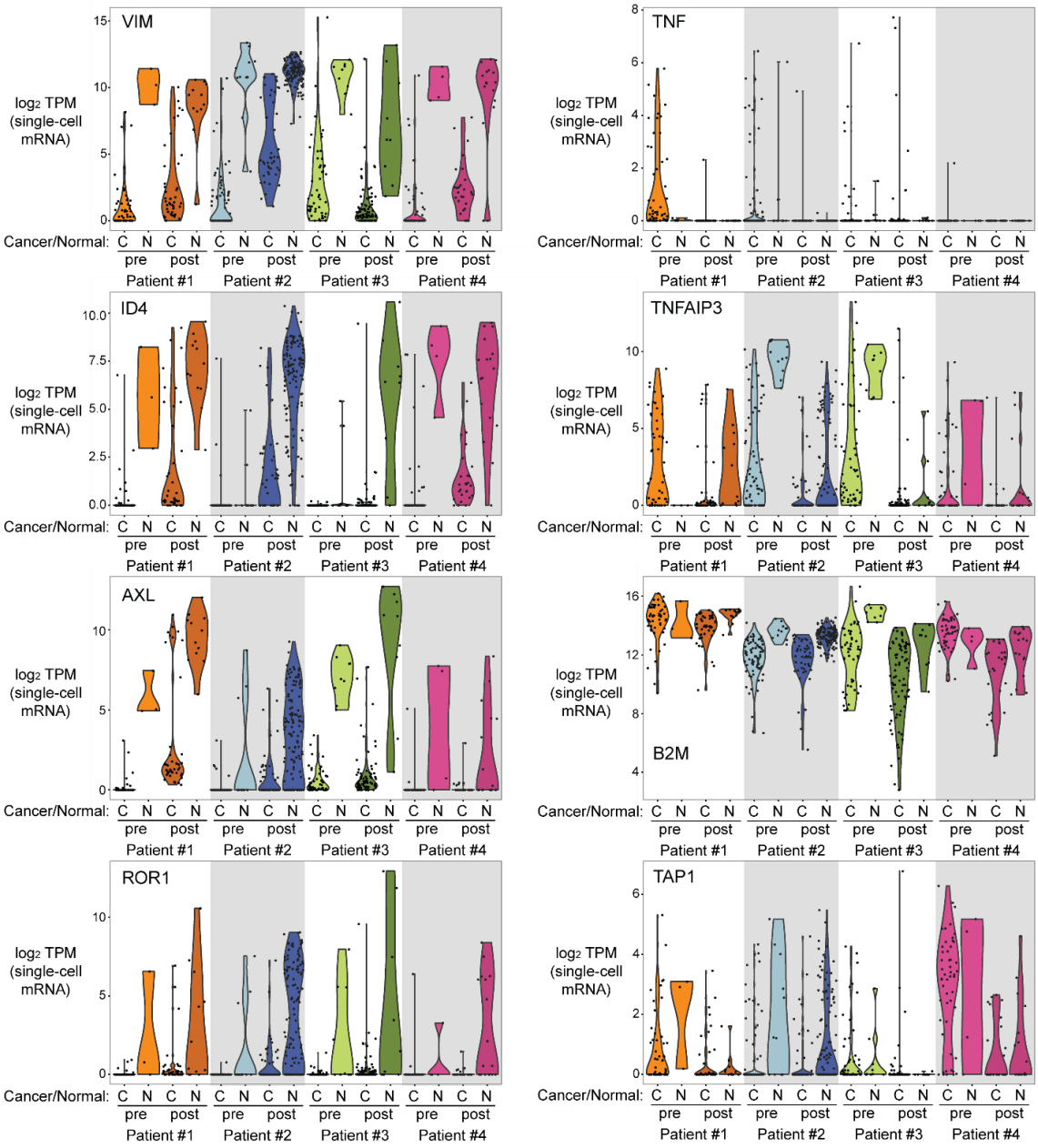
Supplementary Figure 21 | Copy number and subclones inferred from single-cell RNA-Seq of patient #4. scRNA-Seq was performed on pre- and post-treatment cells from patient #4 and copy number was inferred and corroborated with bulk DNA-Seq-based copy for each gene. Subclones were identified using CNVs as described in text and Methods. * indicates CNVs aiding with subclone identification. Subclones assigned (bottom) are defined as in Fig. 2d.



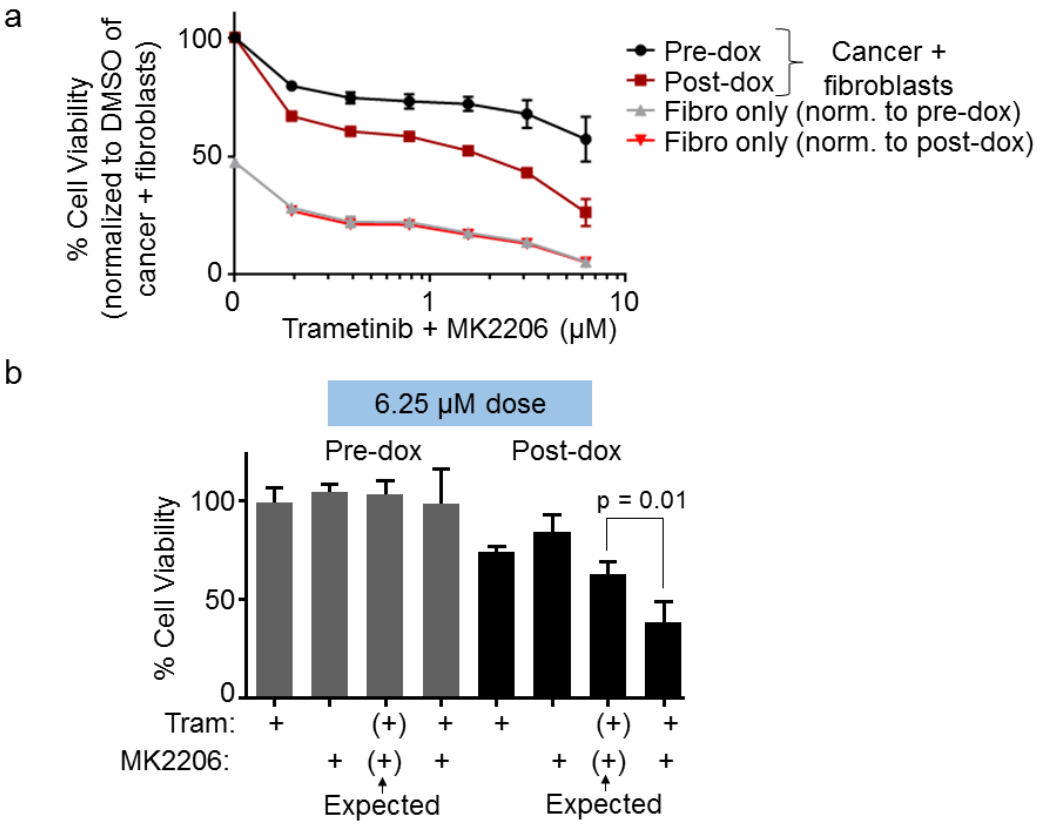
Supplementary Figure 22 | Identification of genes differentially expressed between pre- and post-treatment samples by single-cell RNA-Seq. scRNA-Seq was performed on pre- and post-treatment samples from patients #1-#4 and a two-sample proportion test was performed on each patient for each gene to compare the proportion of cells in pre- vs. post-treatment samples expressing each gene. p-value ranks (x-axis; higher rank indicates higher significances) were plotted against $-\log_{10}$ (p-value) for each gene expressed in at least 1 cell in each patient.



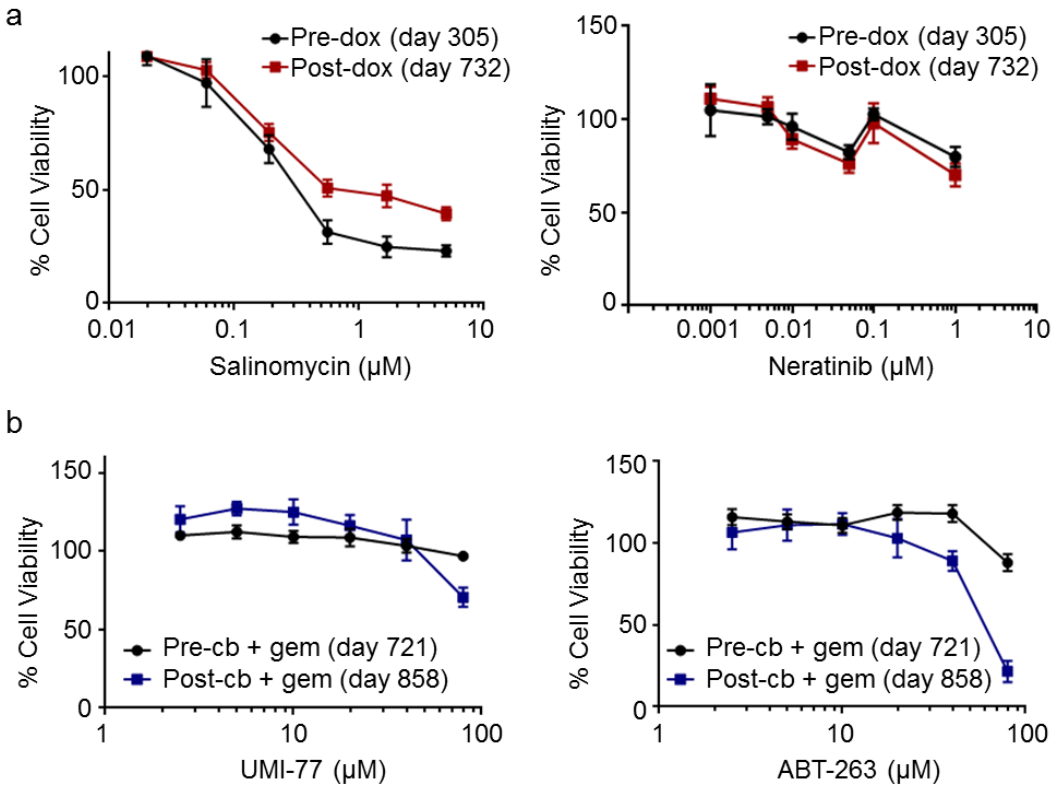
Supplementary Figure 23 | EGFR is increased after treatment in two out of four breast cancer patients by single-cell RNA-Seq. EGFR expression (y-axis) was plotted for each single cell in each patient and a two-sample proportion test was performed to compare the proportion of cells in pre- vs. post-treatment cells expressing EGFR (see p-values). y-axis goes from 0 (left) to 2.0 (right) in each plot.



Supplementary Figure 24 | Detection of EMT, RTK, and immune-avoidance phenotypes in bulk RNA-Seq may be hampered by normal cell contaminants. scRNA-Seq was performed on pre- and post-treatment samples from four patients and cells were identified as normal or cancer based on copy profiles (Supplementary Figs. 18-21). Expression of the 8 indicated genes, found to be differentially expressed between pre- and post-treatment cancer samples, is shown in both normal and cancer populations for each sample using violin plots. Each dot represents a single cell.



Supplementary Figure 25 | Acquired sensitivity to combined MEK and Akt inhibition in post-doxorubicin patient #1 cells. (**a**) Pre- and post-doxorubicin patient #1 cells grown on fibroblasts (or fibroblasts only with no cancer cells) were treated with equimolar doses of indicated drugs and subjected to CellTiter-Glo assay after 3 days of treatment. Data are normalized to DMSO control mean (= 100%) for cancer + fibroblasts wells. Fibroblast-only wells were normalized to pre-dox (grey triangles) or post-dox (red triangles) cancer + fibroblasts DMSO control mean. Dox, liposomal doxorubicin. Cells were from day 305 (pre-dox) or day 732 (post-dox). (**b**) Synergy analysis of 6.25 μM dose of combined trametinib and MK2206 in pre- and post-doxorubicin cells. Expected is by Bliss independence; p-value is by student's t-test. (Fibroblast-only signal was subtracted from cancer + fibroblast readings.)



Supplementary Figure 26 | Lack of acquired sensitivity to salinomycin and neratinib in patient #1 post-doxorubicin cells and to Bcl-2 family inhibition in post-carboplatin cells. (**a**) Cultured pre- (day 305) and post-doxorubicin (day 732) patient #1 cells were treated for 72 hours with indicated drug doses, followed by viability measurement with CellTiter-Glo. Data are normalized to DMSO treatment mean (= 100%). Salinomycin assay was done with Renaissance medium while neratinib assay was done with fibroblast feeder system with fibroblasts withdrawn (see Methods). (**b**) Cultured pre- (day 721) and post-carboplatin + gemcitabine (day 858) cells were treated for 72 hours with indicated drug doses, followed by viability measurement with CellTiter-Glo. Assays were done in Renaissance medium. Made in conjunction with and used by permission of Gajendra Shrestha.

CHAPTER 4

SUMMARY AND PERSPECTIVES

These findings strengthen our understanding of the genetic and phenotypic evolution of breast cancer using an analytical approach of greater depth than any performed previously, to our knowledge, on longitudinal breast cancer samples. They identify specific *ABCB1* fusions and *BRCA2* reversions not previously reported in breast cancer samples, and reveal EMT, RTK activation, and immune-avoidance phenotypes as potential mechanisms providing selective advantage after treatment. Further, these phenotypes likely pre-exist, at least in part, in some pre-treatment subclones, and thus have may have a genetic basis. Drug treatment assays indicate that these phenotypes can suggest drugs to which the cancer has acquired sensitivity, concomitant with their acquired resistance to the current therapy.

Further follow-up on these results may lead to improved treatment options for patients. For example, better approaches are needed to target ABCB1 in drug-resistant cancers. Immune-activating drugs may be effective, in combination with chemotherapy, in preventing acquisition of the immune-avoidance phenotype.

In summary, an improved understanding of the post-treatment acquired genetic and phenotypic changes may lead to more effective treatments for breast cancer, and lead to improved clinical outcomes for patients otherwise having few treatment options.

REFERENCES

1. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144,** 646–674 (2011).

2. Sequist, L. V *et al.* Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci. Transl. Med.* **3,** 75ra26 (2011).

3. Yates, L. R. *et al.* Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat. Med.* **21,** 751–9 (2015).

4. Gottesman, M. M., Fojo, T. & Bates, S. E. Multidrug resistance in cancer: role of ATP-dependent transporters. *Nat. Rev. Cancer* **2,** 48–58 (2002).

5. Antony, V. B. *et al.* Management of malignant pleural effusions. *Eur. Respir. J.* **18,** (2001).

6. Smith, E. & Jayson, G. The current and future management of malignant ascites. *Clin. Oncol.* **15,** 59–72 (2003).

7. Zhang, R. D., Fidler, I. J. & Price, J. E. Relative malignant potential of human breast carcinoma cell lines established from pleural effusions and a brain metastasis. *Invasion Metastasis* **11,** 204–15 (1991).

8. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–20 (2014).

9. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–60 (2009).

10. Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P. & Marth, G. T. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* **27,** 1691–2 (2011).

11. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–303 (2010).

12. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31,** 213–9 (2013).

13. Ramos, A. H. *et al.* Oncotator: cancer variant annotation tool. *Hum. Mutat.* **36,** E2423-9 (2015).

14. Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22,** 568–76 (2012).

15. Chiang, C. *et al.* SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat. Methods* **12,** 966–8 (2015).

16. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin).* **6,** 80–92 (2012).

17. Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30,** 2503–5 (2014).

18. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15,** R84 (2014).

19. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–2 (2010).

20. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19,** 1639–45 (2009).

21. Futreal, P. A. *et al.* A census of human cancer genes. *Nat. Rev. Cancer* **4,** 177–83 (2004).

22. Qiao, Y. *et al.* SubcloneSeeker: a computational framework for reconstructing tumor clone structure for cancer variant interpretation and prioritization. *Genome Biol.* **15,** 443 (2014).

23. Bodenhofer, U., Kothmeier, A. & Hochreiter, S. APCluster: an R package for affinity propagation clustering. *Bioinformatics* **27,** 2463–4 (2011).

24. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–21 (2013).

25. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–11 (2009).

26. Kim, D. & Salzberg, S. L. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol.* **12,** R72 (2011).

27. Futreal, P. A. *et al.* A census of human cancer genes. *Nat. Rev. Cancer* **4,** 177–83 (2004).

28. McDonagh, C. F. *et al.* Antitumor activity of a novel bispecific antibody that targets the ErbB2/ErbB3 oncogenic unit and inhibits heregulin-induced activation

of ErbB3. *Mol. Cancer Ther.* **11,** 582–93 (2012).

29. Robinson, D. R. *et al.* Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat. Genet.* **45,** 1446–1451 (2013).

30. Mullican, S. E. *et al.* Abrogation of nuclear receptors Nr4a3 andNr4a1 leads to development of acute myeloid leukemia. *Nat. Med.* **13,** 730–735 (2007).

31. Houghtaling, S. *et al.* Epithelial cancer in Fanconi anemia complementation group D2 (Fancd2) knockout mice. *Genes Dev.* **17,** 2021–35 (2003).

32. Sakai, W. *et al.* Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers. *Nature* **451,** 1116–1120 (2008).

33. Ellis, M. J. *et al.* Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* **486,** 353 (2012).

34. Tsukamoto, K. *et al.* Allelic loss on chromosome 1p is Associated with progression and lymph node metastasis of primary breast carcinoma. *Cancer* **82,** 317–322 (1998).

35. Ragnarsson, G. *et al.* Loss of heterozygosity at chromosome 1p in different solid human tumours: association with survival. *Br. J. Cancer* **79,** 1468–74 (1999).

36. Youle, R. J. & Strasser, A. The BCL-2 protein family: opposing activities that mediate cell death. *Nat. Rev. Mol. Cell Biol.* **9,** 47–59 (2008).

37. Barber, L. J. *et al.* Secondary mutations in *BRCA2* associated with clinical resistance to a PARP inhibitor. *J. Pathol.* **229,** 422–429 (2013).

38. Aszalos, A. Drug–drug interactions affected by the transporter protein, P-glycoprotein (ABCB1, MDR1): II. Clinical aspects. *Drug Discov. Today* **12,** 838–843 (2007).

39. Calado, R. *et al.* Influence of functional MDR1 gene polymorphisms on P-glycoprotein activity in CD34+ hematopoietic stem cells. *Haematologica* **87,** (2002).

40. Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313,** 1929–35 (2006).

41. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41,** e108 (2013).

42. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–30 (2014).

43. Wang, K. *et al.* MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* **38,** e178 (2010).

44. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).

45. Bullard, J. H., Purdom, E., Hansen, K. D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11,** 94 (2010).

46. Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27,** 1160–7 (2009).

47. Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70 (2012).

48. Luo, J. *et al.* A protocol for rapid generation of recombinant adenoviruses using the AdEasy system. *Nat. Protoc.* **2,** 1236–47 (2007).

49. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28,** 882–3 (2012).

50. Shen, Y. *et al.* ASSIGN: context-specific genomic profiling of multiple heterogeneous biological pathways. *Bioinformatics* **31,** 1745–53 (2015).

51. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344,** 1396–401 (2014).

52. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352,** 189–196 (2016).

53. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161,** 1202–1214 (2015).

54. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14,** 7 (2013).

55. Anastassiou, D. *et al.* Human cancer cells express Slug-based epithelial-mesenchymal transition gene expression signature obtained in vivo. *BMC Cancer* **11,** 529 (2011).

56. Jechlinger, M. *et al.* Expression profiling of epithelial plasticity in tumor progression. *Oncogene* **22,** 7155–69 (2003).

57. Croft, D. *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39,** D691-7 (2011).

58. Kalma, Y., Marash, L., Lamed, Y. & Ginsberg, D. Expression analysis using DNA microarrays demonstrates that E2F-1 up-regulates expression of DNA replication genes including replication protein A2. *Oncogene* **20,** 1379–87 (2001).

59. Kong, L.-J., Chang, J. T., Bild, A. H. & Nevins, J. R. Compensation and specificity of function within the E2F family. *Oncogene* **26,** 321–7 (2007).

60. Czabotar, P. E., Lessene, G., Strasser, A. & Adams, J. M. Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nat. Rev. Mol. Cell Biol.* **15,** 49–63 (2014).

61. Srinivasula, S. M. & Ashwell, J. D. IAPs: what's in a name? *Mol. Cell* **30,** 123–35 (2008).

62. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141,** 1117–34 (2010).

63. Liu, X. *et al.* ROCK inhibitor and feeder cells induce the conditional reprogramming of epithelial cells. *Am. J. Pathol.* **180,** 599–607 (2012).

64. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462,** 108–12 (2009).

65. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27,** 1739–40 (2011).

66. Bild, A. H. *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439,** 353–7 (2006).

67. De Luca, A., Maiello, M. R., D'Alessio, A., Pergameno, M. & Normanno, N. The RAS/RAF/MEK/ERK and the PI3K/AKT signalling pathways: role in cancer pathogenesis and implications for therapeutic approaches. *Expert Opin. Ther. Targets* **16,** S17–S27 (2012).

68. Lombaerts, M. *et al.* E-cadherin transcriptional downregulation by promoter methylation but not mutation is related to epithelial-to-mesenchymal transition in breast cancer cell lines. *Br. J. Cancer* **94,** 661 (2006).

69. Mani, S. A. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133,** 704–15 (2008).

70. Scheurich, P., Thoma, B., Ucer, U. & Pfizenmaier, K. Immunoregulatory activity of recombinant human tumor necrosis factor (TNF)-alpha: induction of TNF receptors on human T cells and TNF-alpha-mediated enhancement of T cell responses. *J. Immunol.* **138,** 1786–90 (1987).

71. Azzawi, M., Hasleton, P. S. & Hutchinson, I. V. TNF-alpha in acute cardiac transplant rejection. *Cytokines. Cell. Mol. Ther.* **5,** 41–9 (1999).

72. Romero, J. M. *et al.* Coordinated downregulation of the antigen presentation machinery and HLA class I/β2-microglobulin complex is responsible for HLA-ABC loss in bladder cancer. *Int. J. Cancer* **113,** 605–610 (2005).

73. Shan, L., Yu, M., Qiu, C. & Snyderwine, E. G. Id4 regulates mammary epithelial cell growth and differentiation and is overexpressed in rat mammary gland carcinomas. *Am. J. Pathol.* **163,** 2495–2502 (2003).

74. de Candia, P., Akram, M., Benezra, R. & Brogi, E. Id4 messenger RNA and estrogen receptor expression: inverse correlation in human normal breast epithelium and carcinoma. *Hum. Pathol.* **37,** 1032–1041 (2006).

75. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141,** 1117–34 (2010).

76. Vereecke, L., Beyaert, R. & van Loo, G. The ubiquitin-editing enzyme A20 (TNFAIP3) is a central regulator of immunopathology. *Trends Immunol.* **30,** 383–391 (2009).

77. Kaer, L. Van, Ashton-Rickardt, P. G., Ploegh, H. L. & Tonegawa, S. TAP1 mutant mice are deficient in antigen presentation, surface class I molecules, and CD4−8+ T cells. *Cell* **71,** 1205–1214 (1992).

78. Serreze, D. V, Leiter, E. H., Christianson, G. J., Greiner, D. & Roopenian, D. C. Major histocompatibility complex class I-deficient NOD-B2m-null mice are diabetes and insulitis resistant. *Diabetes* **43,** (1994).

79. Xu, J., Lamouille, S. & Derynck, R. TGF-β-induced epithelial to mesenchymal transition. *Cell Res.* **19,** 156–172 (2009).

80. Sukumar, S. *et al.* Compromised HOXA5 function can limit p53 expression in human breast tumours. *Nature* **405,** 974–978 (2000).

81. Falchook, G. S. *et al.* Activity of the oral MEK inhibitor trametinib in patients with advanced melanoma: a phase 1 dose-escalation trial. *Lancet. Oncol.* **13,** 782–9 (2012).

82. Hirai, H. *et al.* MK-2206, an allosteric Akt inhibitor, enhances antitumor efficacy by standard chemotherapeutic agents or molecular targeted drugs in vitro and in vivo. *Mol. Cancer Ther.* **9,** 1956–67 (2010).

83. Wong, K.-K. *et al.* A phase I study with neratinib (HKI-272), an irreversible pan ErbB receptor tyrosine kinase inhibitor, in patients with solid tumors. *Clin. Cancer Res.* **15,** 2552–8 (2009).

84. Gupta, P. B. *et al.* Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* **138,** 645–59 (2009).

85. Chappell, W. H. *et al.* Ras/Raf/MEK/ERK and PI3K/PTEN/Akt/mTOR inhibitors: rationale and importance to inhibiting these pathways in human health. *Oncotarget* **2,** 135–64 (2011).

86. Serra, V. *et al.* PI3K inhibition results in enhanced HER signaling and acquired ERK dependency in HER2-overexpressing breast cancer. *Oncogene* **30,** 2547–57 (2011).

87. Todd, J. R., Scurr, L. L., Becker, T. M., Kefford, R. F. & Rizos, H. The MAPK pathway functions as a redundant survival signal that reinforces the PI3K cascade in c-Kit mutant melanoma. *Oncogene* **33,** 236–45 (2014).